New Curve Fitting Method for Petroleum Engineering Applications

By

Muhammad Aizuddin bin Mohd Rosli

13653

Dissertation submitted in partial fulfillment of

the requirements for the

Bachelor of Engineering (Hons)

(Petroleum)

SEPTEMBER 2014

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF ORIGINALITY

The author hereby declares that the contents of his submitted thesis are free from any material already published by another author nor does it contain statements lifted without due acknowledgement of the sources. He similarly attests that materials taken from other sources are properly quoted.

Thus, except those which have been duly acknowledged, recognized and quoted in the text, the content of this thesis has been authentically produced by the author himself though he may have received assistance from others on style, presentation and language expression.

(Muhammad Aizuddin bin Mohd Rosli)

CERTIFICATION OF APPROVAL

New Curve Fitting Method for Petroleum Engineering Applications

By

Muhammad Aizuddin bin Mohd Rosli

A project dissertation submitted to the
Petroleum Engineering Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
BACHELOR OF ENGINEERING (Hons)
(PETROLEUM ENGINEERING)

Approved by,

(Samsul Ariffin bin Abdul Karim)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
SEPTEMBER 2014

ABSTRACT

Data collected in petroleum engineering industry may come in a non-distribution form. Usually the data is really important for the approximation and future prediction purposes. It may lead to the increasing of time and cost if wrong prediction happened. So, to avoid the problems that may occurred, the best fit is needed to represent the data. Curve fitting is the process of creating a curve or mathematical function that will produce the best fit to a series of data points. It usually uses to design and control the shape of complex curves and surfaces. There are several of methods of approximation has been introduced such are Polynomial, Gaussian, Sine, Fourier Series and Smoothing Spline but unfortunately not yet meet the high expectations. For example, Polynomial approach in the algorithm which contains strongly oscillating properties and huge number of arithmetical operations for computations. Unsuitable type of curve fitting may destroy the original feature of the data. This research is about the invention of new curve fitting method that combined two interpolation method which are Cubic Spline and Piecewise Cubic Hermite Interpolation Polynomial (PCHIP) with the reduction of data method. This research will show the comparison between the new curve fitting method and the existing method by applying it to several sets of petroleum engineering data.

ACKNOWLEDGMENT

First and foremost, all thanks and praise to Allah the almighty for the guidance and blessing for giving the strength to pursue the research for this final year project. In this opportunity, I would like to convey my deepest appreciation and gratitude to my supervisor, Mr Samsul Ariffin bin Abdul Karim for his guidance, advice and knowledge throughout the duration of the project. Without his support and encouragement, I could not harvest experiences and knowledge as much that I wanted.

Besides, the appreciation will not complete without thanking my co-supervisor, Izzatullah Muhammad from Emerson Process Management Kuala Lumpur for his assistance and advices. I also would like to thank the Petroleum Engineering Department of Universiti Teknologi Petronas, the lecturers and the coordinators of the final year project for their readiness in providing full support and co-operation during the project period

Last but not least I, I like to take this opportunity to express my appreciation to my family, especially to my parent for giving me full support to complete this industrial training. May ALLAH SWT bless them all and thank you so much for their support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

The project is related to the understanding and applications about the interpolation and curve fitting methods used as an approximation and optimization tools in petroleum engineering field. Splines are being use widely nowadays. The application of spline method could be observed in the most sophisticated petroleum engineering software nowadays such as Schlumberger Eclipse, CMG, Tempest, etc. Besides, spline is being applied widely especially in the well placement optimization and history matching alongside with other mathematical algorithms i.e. genetic algorithm and gradient based algorithm. This could be proven by Lee at el (1986) where they used splines for history matching purposes in single-phase two-dimensional reservoirs. The main objective of this work is to investigate the use of cubic spline and Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolation and curve fitting methods in petroleum engineering data by proposing a new curve fitting involving data reduction that outperform the existing method .To achieve this objective, this work will be conducted with assistance of mathematical simulation software, Matlab.

1.2 PROBLEM STATEMENT

Petroleum engineering data set is usually came in sparse distributed order which would be in non-uniform distributions and in order to obtain the best reservoir performance results, specific equations and algorithm are needed as a tool for the approximation technique (Mehdi, 2011). Various methods of approximation had been introduced to approach the best approximation in petroleum engineering performance but unfortunately not yet meet the high expectations. One of the weaknesses of those methods are they are using polynomials approach in the algorithm which contains strongly oscillating properties and huge number of arithmetical operations for computation. It may destroy the features of the original data (Karim, 2013). By choosing the unsuitable type of curve fitting will destroy the original feature of the data. So the best curve fitting method needed in process to obtain the best fit that will represent the trend of data and for approximation purpose. Different types of interpolation and curve fitting methods will be approached in this project to overcome the weakness and as a part for constructing a new curve fitting method.

1.3 OBJECTIVES

1. To investigate the existing methods for data interpolation and data fitting such as Polynomial, Gaussian, Fourier, Sine and Spline Smoothing.
2. To propose a new curve fitting method to overcome the weakness of existing curve fitting methods. Our new curve fitting method are efficient and reliable with less error and data reduction.
3. To develop new algorithm based on new curve fitting method.
4. To compare the performance between new curve fitting method with the existing methods.

1.4 SCOPE OF STUDY

In this project, different types of interpolation and curve fitting method will be applied to the petroleum engineering data. All the data set will be analyzed by using all the different methods and the result will be compared to produce a smooth curve with piecewise polynomial attribute. Based on the best fit interpolation and curve fitting, the high integrity of well performance could be analyzed and the best results could be presented. A new data fitting method also will be applied in this project by proposing new algorithm involving data reduction

1.5 THE RELEVANCY OF THE PROJECT

This project is relevant to be conducted because data collected in petroleum engineering industry always come in non-uniform distribution and it is difficult to obtain the best fit to represent the data trend. Because of that matter, additional time and cost are needed to overcome the problems that may occur. Thus, it will lead to the loss in profit for the operator. In petroleum engineering whether in reservoir and drilling part, correct planning is needed to give the best performance result. So, the best estimation and approximation of the planning data are needed to avoid any problem occur during operation and give the best results.

1.6 FEASIBILITY OF THE PROJECT WITHIN THE SCOPE AND TIME FRAME

This project is feasible and can be finished within the time frame because it only needs several analyses from the real drilling report, internet articles and some opinion from the engineer in the petroleum engineering area. Most of this project contains are being conducted using the MATLAB software and it may takes about 2 to 3 months to be completed.

CHAPTER 2

LITERATURE REVIEW

In this section, three different parts of review will be made which are the interpolation, curve fitting and the new curve fitting method. All steps and equations will be explained in more detailed for every part.

2.1 INTRODUCTION OF INTERPOLATION CURVE

One of the mathematical representations that help user to design and control the shape of complex curves and surfaces is a spline curve. It can be divided into two different types of curves which are interpolating curve and approximation curve. Interpolating curve passes through each control point while approximation curve passes near to the control points but not necessarily through them

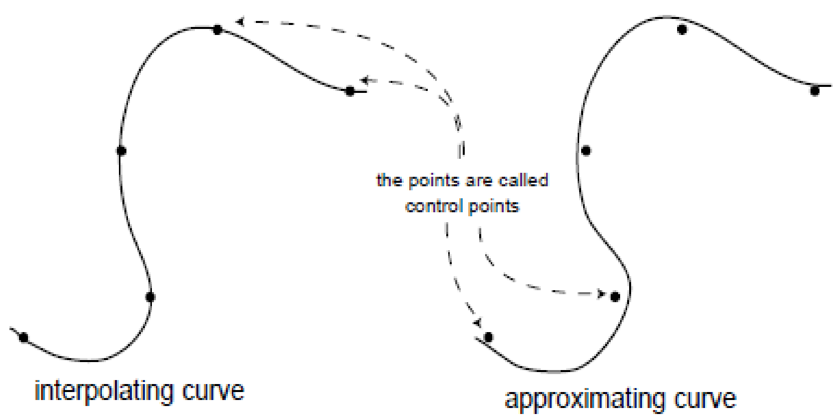


Figure 1: Interpolating and Approximation Curve

2.2 OVERVIEW OF APPROXIMATION AND INTERPOLATION THEORY

Smooth and continuous curve to pass through or closer to a set of data points can be found by using the curve fitting method. Curve fitting method could indicate the best trend of data set (Won et al, 2005) without need to go through all the data points.

Two approaches can be done to fit the curve of a set of data points. The first one can be called as a Collocation, where every data point will be passed through by the curve. This can be used either when the data is accurate or the data are known to be generated by evaluating a complicated function at a discrete set of points (Singiresu, 2002).

There are three type of functions can be used to approximate a set of data points which are Polynomial, trigonometric or exponential. Piecewise curve fitting is used where a specified function is made through sub-groups of data points in some cases. Another approach is where the curve is made to represent the behavior of the data. According to Singiresu (2002), this approach is convenient to be used when the number of unknown coefficients is less than the data points or when the data may have being disturbed by some noise or contain some errors. Kruger (n.d.) stated that the value of a function between data points can be estimated by using interpolation method without knowing the actual function.

According to Henrici. (1982), as cited in Kruger (n.d.), there are two major categories for the interpolation methods. First is the global interpolation where single equation that fits all the data point needs to be constructed and the degree of the polynomial equation mostly high. Even though smooth curve will be the result from this interpolation, it may lead to severe oscillation and it is not suitable to be applied in engineering applications.

The other category is the piecewise interpolation. Low degree of polynomial function will be constructed in this method. First degree polynomial can be known as a linear interpolation. Second degree polynomial is called as quadratic and lastly, the third degree polynomial is cubic spline. The curve will be smoother with the increasing of the degree of spline.

For a given set of N+1 data points (x_0, y_0) , (x_1, y_1),…..(x_n, y_n), the objective is to find the N^{th} order polynomial function that can match them. But, the unknown coefficients and equations will be increasing as the number of the data points are increases. According to Singiresu (2002), the polynomial will tend to increase acutely due to the errors as the order n is increasing. Singiresu (2002) stated that the high degree of polynomial often lead to the unnecessary oscillations and wiggles. Thus, the polynomial interpolation will not give an accurate interpretation to the given data. Therefore spline interpolation is the most suitable to be use becauselarge data point can be used and the data behavior can be maintained. There are three common spline interpolations used which are linear, quadratic, and cubic splines. Cubic spline is the most recommended if want to obtain a smoother curve. Figure below show the example of high order polynomial interpolation that show the wildly oscillations.

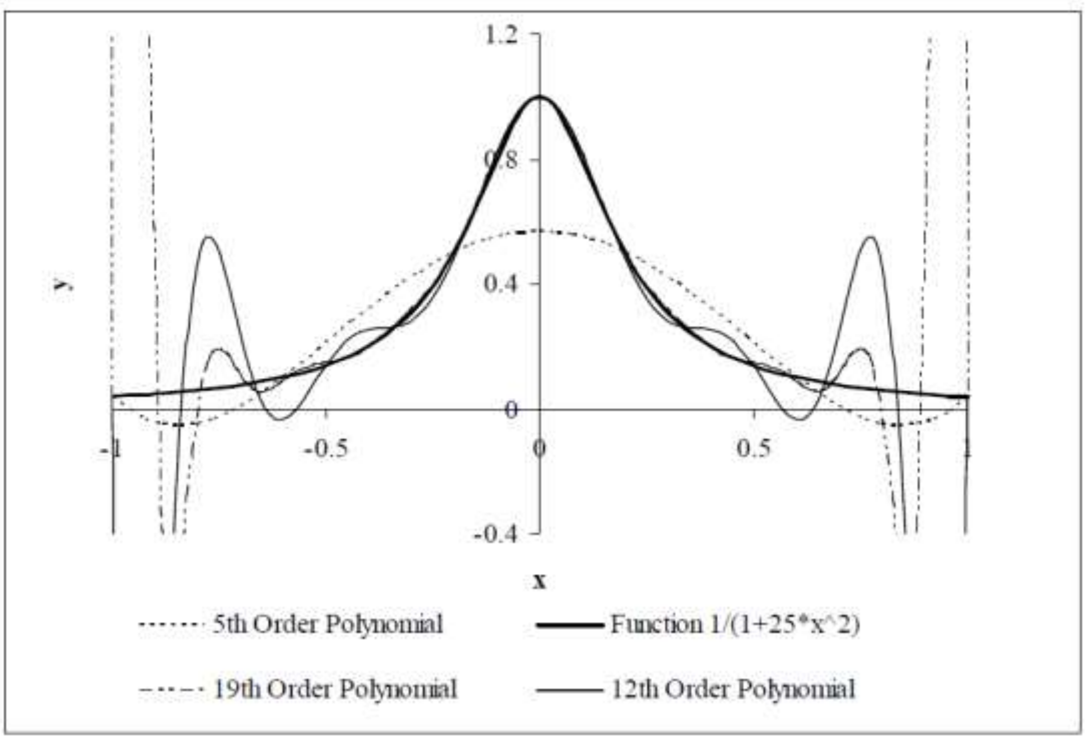


Figure 2: High Order Polynomial Interpolation

Thus, we may conclude that interpolation of the data by using the higher degree of polynomial may be resulting the interpolation oscillates wildly. This behavior is unacceptable and cubic spline interpolation can be used to overcome this problem. This is because cubic splines may avoid the wild oscillation as well as the overshoot and the obtained interpolating curves are acceptable or desirable as desired by the user or respective designer (Kruger, n.d.).

2.2.1 Linear Interpolation

According to Singiresu (2002), linear spline represents a straight line between the data points (knots). Let $n+1$ data points be available as $[x_i, f(x)]$, $i = 0,1,2,...n.$ Singiresu (2002) explained that between two data points next to each other which is $[x_{i-1,} f(x_{i-1})]$ and $[x_{i,} f(x_i)]$, the equation of the line that join the two points is defined in Equation (1) where the function $f_i(x)$ represent a set of n piecewise linear equation (splines) using the $n+1$ data points and Equation (2) is the slope between x_{i-1} and x_i.

According to Addendum (n.d.), there is no different between linear spline interpolation and linear polynomial interpolation. The author stated that linear splines still using the data from the two consecutive data points. Also at the interior points of the data, the slope will changes abruptly which means that the first derivative is not continuous at these points.

$$f_i = f(x_{i-1}) + \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}(x - x_{i-1}) \; ; i = 1,2,...n. \tag{1}$$

$$\frac{f(x_i)-f(x_{i-1})}{x_i-x_{i-1}} \tag{2}$$

2.2.2 Quadratic Spline

A type of spline where the second order polynomial approximates the data between two consecutive data points or knots. According to Singiresu (2002), the quadratic spline that given $x_i, f(x_i), i=0,1,2,....,n$ to denote $n+1$ data points, the equation of the quadratic spline between the data points $x_{i-1}, f(\mathrm{x}_{i-1})$ and $x_i, f(\mathrm{x}_i)$ can be expressed in Equation (2) where a_i, b_i and c_i are the unknown coefficients.

$$f_i(x)=a_i+b_ix+c_ix^2; i=1,2,...,n \tag{3}$$

3*n* coefficients need to be evaluated for the n intervals. To calculate the 3*n* unknowns, 3*n* equations need to be build and will be solved simultaneously. According to Singiresu (2002), these 3*n* equations can be found by the following conditions.

The function value at the interior knot x_i must be equal to $f(x_i)$ whether it is computed using $f(x_i)$ or $f(x_{i+1})$ which is

$$f_i(x_i)=a_i+b_ix_i+c_ix_i^2 = f(x_i); i=1,2,......n, \tag{4}$$

and

$$f_{i+1}(x_i)=a_{i+1}+b_{i+1}x_i+c_{i+1}x_i^2 = f(x_i); i=1,2,......n-1, \tag{5}$$

Equations (3) and (4) gives ($2n-2$) conditions.

The end points x_0 and x_n must be passed by first and last functions, $f_1(x)$ and $f_n(x)$.

$$f_1(x_0) = a_1 + b_1 x_0 + c_1 x_0^2 = f(x_0), \tag{6}$$

and

$$f_n(x_n) = a_n + b_n x_n + c_n x_n^2 = f(x_n). \tag{7}$$

The slope of two quadratic spline or first derivative is continuous at the interior points. For example, the derivative of Equation (5) gives the slope as

$$f_i^{'}(x) = b_i + 2x_i c_i ; i = 1, 2, ..., n \tag{8}$$

and hence, the continuity of slope leads to

$$f_i^{'}(x_i) = f_{i+1}^{'}(x_{i+1}), \tag{9}$$

that is

$$b_i + 2c_i x_i = b_{i+1} + 2c_{i+1} x_i ; i = 1, 2, ...n - 1. \tag{10}$$

Equation above give (n- 1) conditions.

Thus, the total number of equations is (3n - 1) equations. One more equation is needed to calculate all $3n$ constants. Several possible conditions can be used for the calculations. For example, at final point (x_n) the second derivative can be assumed to be zero.

$$f_n^{''}(x_n) = 2c_n = 0, \text{or}\; c_n = 0. \tag{11}$$

2.2.3 Cubic Spline

Won et al. (2005) stated in his thesis that the smoothness of the quadratic curve is not smooth enough since the second-order derivatives of quadratic polynomial for adjacent subintervals cannot be made to conform for each other. However, cubic splines have proven its goodness in term of complexity and accuracy. A cubic spline through a set of data points is a curve obtained by joining each point to the next with a cubic polynomial, where the important part is that the adjoining cubic spline must a have matching first and second derivative at their common point. The equation of the cubic spline in the i^{th} interval $[x_{i-1}, x_i]$, is show in Equation (12).

$$f_i(x) = a_i + b_i x + c_i x^2 + d_i x^3; i = 1, 2, ...n, \tag{12}$$

where a_i, b_i, c_i, and d_i is the 4n coefficients for $i = 1, 2, ...n,$

According to Hoffman (2001) these 4n coefficients a_i, b_i, c_i and d_i can be evaluated by using the following condition:

1. The function values $f_i(x) = f(i = 2, 3......n)$, must be the same in the two splines on either side of x_i at all of the n-1 interior points. This constraint yields 2(n-1) conditions.

2. The slope of the first order derivative of the two splines on either side of point x_i must be equal at all of the n-1 interior points.

$$f_i'(x_i) = f_{i+1}'(x_i). \tag{13}$$

This constraint yields (n-1) conditions.

3. The second derivative of the two splines on either side of point x_i must be equal at all of the $n-1$ interior points. This constraint yields (n-1) conditions.

4. The first and last spline must pass through the first and last points. This shown in Equation (14).

$$f_1(x_1) = f_1 \text{ ,}$$
$$f_n(x_{n+1}) = f_{n+1} \, . \tag{14}$$

This yield two conditions.

5. To obtain unique polynomial, two more equations are needed. Several types of conditions can be used to obtain the two equations;

   - For this project, only natural spline will be used in which set the two conditions as in Equation (15).

$$f''(x_0) = 0, \; f''(x_n) = 0 \tag{15}$$

Thus, the curvature must be specified at the first (x_1) and last (x_{n+1})points as shown in Equation (16).

$$f_1''(x_1) = f_1'' \text{ and } f''(x_{n+1}) = \mathrm{f}_{n-1}'' \, . \tag{16}$$

This constraint yields 2 conditions.

2.3 Computing Cubic Spline Interpolation

To fit a different interpolation coefficient, cubic spline interpolation to each successive pair of points, we will define a set of functions $S_i(x)$, on each interval $[x_i, x_{i+1}]$.To keep the solution smooth we would like the match the first and second derivatives, as well as the function itself, at each grid point. The conditions are

$$f_i(x_i) = f_i \tag{17}$$

$$f_i(x_{i+1}) = f_{i+1}(x_{i+1}) \tag{18}$$

$$f_i^{'}(x_{i+1}) = f_{i+1}^{'}(x_{i+1}) \tag{19}$$

$$f_i^{''}(x_{i+1}) = f_{i+1}^{''}(x_{i+1}) \tag{20}$$

Equation (17) until (20) gives a total of 4-2n conditions. As stated before, for cubic spline interpolation, we need $4n$ parameters to be determined. So, we lack of two more condition $4n$-2 < $4n$. We can uniquely determine a set of cubic spline by adding two additional conditions. There exists many type of additional condition somehow, but as stated before natural boundary conditions will be used.

$$f_0^{''}(x_0) = f_{n-1}^{''}(x_n) = 0 \tag{21}$$

General equation of cubic spline is given as follows:

$$f_i(x) = a_i + b(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \tag{22}$$

Substituting Equation (17) into Equation (22), to find the value a_i

$$a_i = f_i \tag{23}$$

Next, from Equation (18), where $h = x_{i+1} - x_i$

$$a_i + b_i h_i + c_i h_i^2 + d_i h_i^3 = a_{i+1} \tag{24}$$

Derive Equation (22) into 1st and 2nd order differentiation, will give the following:

$$f_i^{'}(x) = b_i + 2c_i(x - x_i) + 3d_i(x - x_i)^2 \tag{25}$$

$$f_i^{''}(x) = 2c_i + 6d_i(x - x_i) \tag{26}$$

Then we substitute Equation (19) and (20) into Equation (25) and (26)

$$b_i + 2c_i h_i + 3d_i h_i^2 = b_{i+1} \tag{27}$$

$$2c_i + 6d_i h_i = 2c_{i+1} \tag{28}$$

As stated before, we lack of two additional conditions, so first we define one additional number.

$$c_n = c_{n-1} + 3d_{n-1}h_{n-1} \tag{29}$$

To obtain the parameter d_i, substitute Equation (21) into Equation (26).

$$c_0 = 0 \text{ and } 2c_n = 2c_{i-1} + 6d_{n-1}h_{n-1} = 0 \tag{30}$$

Rearranging the Equation (30), we get the equation for d_i.

$$d_i = \frac{c_{i+1} - c_i}{3h_i} \tag{31}$$

Thus we now have determined all of the a_i and d_i are fully determined by the c_i.

Next, to get b_i,substitute Equation (31) into Equation(24)

$$a_i + b_i h_i + c_i h_i^2 + \frac{c_{i+1} - c_i}{3h_i} h_i^3 = a_{i+1} \tag{32}$$

Rearrange Equation (31), we obtained the following equation for b_i.

$$b_i = \frac{a_{i+1} - a_i}{h_i} - \frac{h_i}{3}(2c_i + c_{i+1}) \tag{33}$$

Finally, to obtain the parameter of c_i, first we reduce the index of Equation (17) by 1.

$$b_{i-1} = \frac{a_i - a_{i-1}}{h_{i-1}} - \frac{h_{i-1}}{3}(2c_{i-1} + c_i) \tag{34}$$

Then, substitute Equation (31) into Equation (27).

$$b_{i+1} = b_i + 2c_i h_i + (c_{i+1} - c_i)h_i = b_i + h_i(c_i + c_{i+1}) \tag{35}$$

Again we reduce the index by 1

$$b_i = b_{i-1} + h_{i-1}(c_{i-1} + c_i) \tag{36}$$

Substitute Equation (34) into the right hand side of Equation (36) and Equation (17) into the left hand side of Equation (20).

$$\frac{a_{i+1} - a_i}{h_i} - \frac{h_i}{3}(2c_i + c_{i+1}) = \frac{a_i - a_{i-1}}{h_{i-1}} - \frac{h_{i-1}}{3}(2c_{i-1} + c_i) + h_{i-1}(c_{i-1} + c_i) \tag{37}$$

Rearrange Equation (37)

$$3\frac{a_{i+1} - a_i}{h_i} - 3\frac{a_i - a_{i-1}}{h_{i-1}} = h_i c_{i+1} + 2c_i(h_i - h_{i-1}) + c_{i-1}h_{i-1} \tag{38}$$

This equation is defined for i=0……,n-1. We can extend it to include an extra parameter a_n that we actually not interested in, to give

$$\begin{pmatrix} 1 & 0 & \cdots & & & 0 \\ h_0 & 2(h_0+h_1) & h_1 & & & \vdots \\ 0 & h_1 & 2(h_1+h_2) & h_2 & & \\ \vdots & & \ddots & \ddots & \ddots & \\ 0 & & 0 & h_{n-2} & 2(h_{n-1}+h_{n-2}) & h_{n-1} \\ 0 & & & & 0 & 1 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ \\ c_n \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{3}{h_1}(a_2 - a_1) - \frac{3}{h_0}(a_1 - a_0) \\ \vdots \\ \frac{3}{h_{n-1}}(a_n - a_{n-1}) - \frac{3}{h_{n-2}}(a_{n-1} - a_{n-2}) \\ 0 \end{pmatrix} \tag{39}$$

If the grid points are equally spaced with $h_i = h$ for some number h, then

$$\begin{pmatrix} 1 & 0 & \cdots & & & 0 \\ h & 4h & h & 0 & & \vdots \\ 0 & h & 4h & h & 0 & \\ \vdots & & \ddots & \ddots & \ddots & \\ 0 & & 0 & h & 4h & h \\ 0 & & & & 0 & 1 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ \\ c_n \end{pmatrix} = \frac{3}{h} \begin{pmatrix} 0 \\ a_0 - 2a_1 + a_2 \\ \vdots \\ \vdots \\ a_{n-2} - 2a_{n-1} + a_n \\ 0 \end{pmatrix} \tag{40}$$

The linear system of Equations in (39) and (40) are triadiagonal and positive definite. Thus the solution will unique and LU decomposition, Thomas algorithm or Gauss Jordan can be used to solve the linear equation.

2.3.1 Computing Cubic Spline Interpolating Example

Given data below is a simple data of car velocity versus time when the car is accelerating.

Table 1: Car Accelrating

Time,s	Velocity, m.p.h.
2	25
3	36
4	52
5	59

1. First we find a_i by using Equation (23).

   a_o=25 , a_1=36 , a_2=52 , a_3=59

2. Next by using Equation (40), we can find the parameter of c since the value of h is equally spaced ($h=1$). The right hand side is equal to the linear system of equation (in matrix form)

   So, it becomes,

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 15 \\ -27 \\ 0 \end{pmatrix} \tag{41}$$

   Multiplying the matrices on the left and setting like components equal give the equivalent system of equations, By solving the system of linear, we obtain the value of c.

$$c_o = 0, c_1 = 5.8, c_2 = -8.2, c_3 = 0 \tag{42}$$

3. To calculate the value of parameter b, we use Equation (33) with h=1

$$\begin{aligned} b_o &= \frac{(36-35)}{1} - \frac{1}{3}(2(0)+5.8) \\ &= 9.07 \end{aligned}$$

$$\begin{aligned} b_1 &= \frac{(52-36)}{1} - \frac{1}{3}(2(5.8)+(-8.2) \\ &= 14.87 \end{aligned}$$

$$\begin{aligned} b_2 &= \frac{(59-52)}{1} - \frac{1}{3}(2(-8.2)-0) \\ &= 12.47 \end{aligned} \tag{43}$$

4. Next, for the value of parameter d, we use Equation (31).

$$\begin{aligned} d_0 &= \frac{5.8-0}{3} \\ &= 1.93 \end{aligned}$$

$$\begin{aligned} d_1 &= \frac{(-8.2-5.8)}{3} \\ &= -4.67 \end{aligned}$$

$$\begin{aligned} d_2 &= \frac{(0-(-8.2)}{3} \\ &= 2.73 \end{aligned} \tag{44}$$

5. Combine Equation (41), (42), (43) and (44) into general equation of cubic spline.

$$f_0(x) = a_0 + b_0(x - x_0) + c_0(x - x_0)^2 + d_0(x - x_0)^3$$
$$= 25 + 9.07(x-2) + 0(x-2)^2 + 1.93(x-2)^3$$
$$= 1.93x^3 - 11.58x^2 + 32.23x - 8.58$$

$$f_1(x) = a_1 + b_1(x - x_1) + c_1(x - x_1)^2 + d_1(x - x_1)^3$$
$$= 36 + 14.87(x-3) + 5.8(x-3)^2 + (-4.67)(x-3)^3$$
$$= -4.67x^3 + 47.83x^2 - 146.02x - 169.68$$

$$f_2(x) = a_2 + b_2(x - x_2) + c_2(x - x_2)^2 + d_2(x - x_2)^3$$
$$= 52 + 12.47(x-4) + (-8.2)(x-4)^2 + 2.73(x-4)^3$$
$$= 2.73x^3 - 40.96x^2 + 209.11x - 303.80$$

Thus, $$f(x) = \begin{cases} 1.93x^3 - 11.58x^2 + 32.23x - 8.58, [2,3] \\ -4.67x^3 + 47.83x^2 - 146.02x - 169.68, [3,4] \\ 2.73x^3 - 40.96x^2 + 209.11x - 303.80 [4,5] \end{cases} \tag{45}$$

2.3.2 Matlab Implementation

The main objective of this section is to show the use of cubic spline interpolation by using matlab scheme. In Matlab, the cubic spline can be applied by using the command line: interp1 (x,y,'spline')

2.3.2.1 Example

```
%Sample matlab coding for data interpolation by using cubic spline
x=2:5;
y=[25 36 52 59];
x1=2:0.01:5;
y1=interp1(x,y,x1,'spline');
plot(x,y,'o',x1,y1,'-')
```

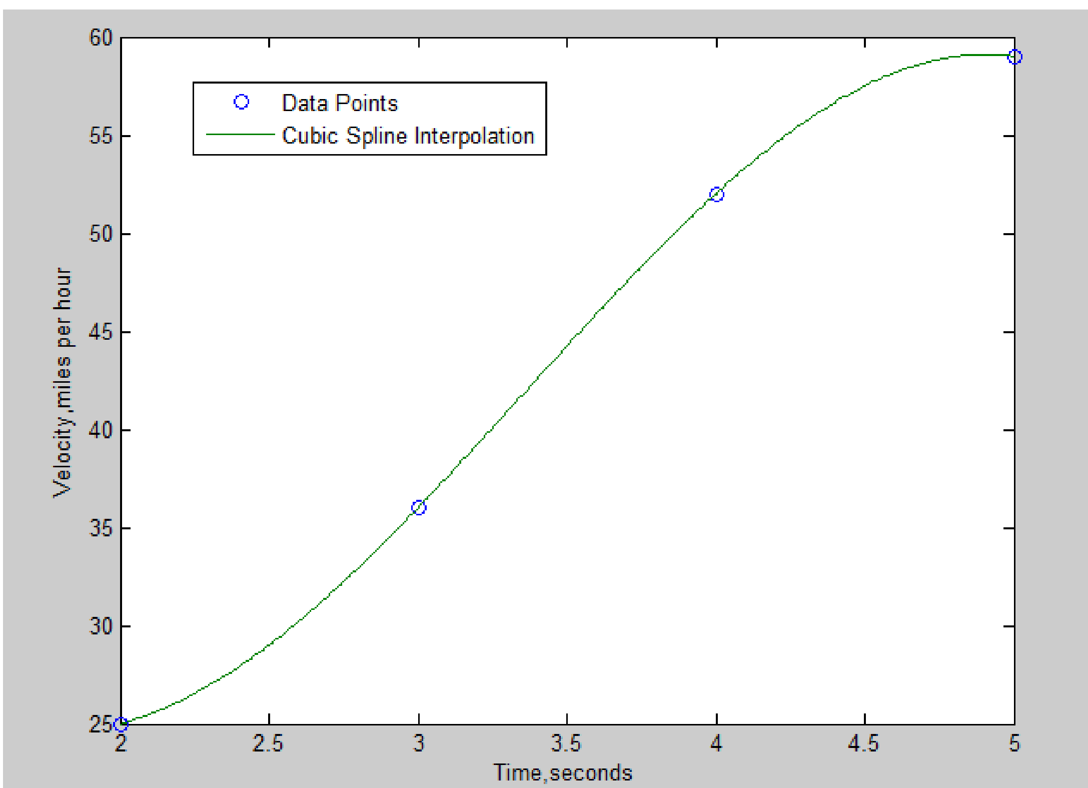
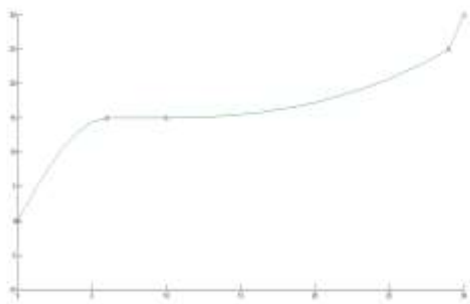
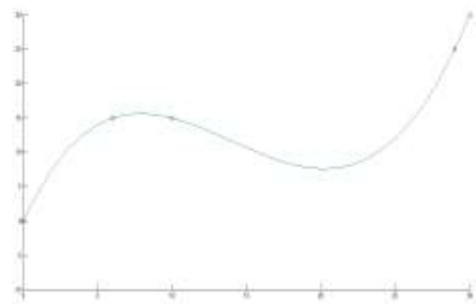


Figure 3: Cubic Spline Interpolation Car Accelerating

2.3.3 Comparison between Piecewise Cubic Hermite Interpolating Polynomial and Cubic Spline Interpolation

Table 2: Comparison between PCHIP and Cubic Spline Interpolation

PCHIP Interpolation	**Cubic Spline Interpolation**
	
• Can be used for monotonicity-, positivity- and convexity-preserving data interpolation	• Can be used to the existing method such as linear regression and non-linear regression
• C^1 continuity.	• C^2 continuity
• Curves tend to overshoot	• Smooth interpolating Curves

2.4 INTRODUCTION TO CURVE FITTING

Curve fitting is the process of creating a curve, or mathematical function, that produce the best fit to a series of data points and it will possibly subject to constraints. Interpolation can also be involved in the curve fitting, where an exact fit to the data is needed, or smoothing, in which a "smooth" function is constructed for the approximation of data. Another topic that is related to curve fitting is regression analysis where it will focuses more on statistical interference. For example, a calculation of the value of uncertainty that present in a curve fitting with random error. The data fitting method is it can be used as an aid for data visualization, i.e. to create values of a function where no data provided and to sum up between two or more variables relationship. Several types of curve fitting that will be discussed in this project are as follows:

(1) Polynomial

(2) Gaussian

(3) Fourier series

(4) Sine

(5)Spline Smoothing.

2.4.1 Linear Regression

This type of fitting method is known as the first degree of polynomial fitting. Below is the example on how this method is being applied.

Let a straight line $\mathrm{y} = a + bx$ (46)

Which is fitted to the data points $(x_{1,}\mathrm{y}_1),(x_{2,}\mathrm{y}_2),(x_{3,}\mathrm{y}_3),.........,(x_{n,}\mathrm{y}_n),$

Let y_{λ_1} be the theoretical value for x_1 then the error, $e_1 = \mathrm{y}_1 - \mathrm{y}_{\lambda_1}$
Then,

$$e_1 = \mathrm{y}_1 - (a + bx_1)$$

$$e_1^2 = (\mathrm{y}_1 - a + bx_1)^2 \tag{47}$$

Now we have the sum of square error, S

$$S = e_1^2 + e_2^2 + e_3^2 + ...+ e_n^2$$

$$S = \sum_{i=1}^{n} e_i^2$$

$$S = \sum_{i=1}^{n} (\mathrm{y}_i - a - bx_i)^2 \tag{48}$$

By using the principal of least squares, the value of S is minimum, thus,

$$\frac{\partial S}{\partial a} = 0 \tag{49}$$

and

$$\frac{\partial S}{\partial b} = 0 \tag{50}$$

Solve equations (48) and (49), and dropping the suffix, we obtain equation below

$$\sum \mathrm{y} = na + b\sum x \tag{51}$$

$$\sum \mathrm{xy} = a\sum x + b\sum x^2 \tag{52}$$

The equation (50) and (51) are known as normal equations. By solving equations (50) and (51), we get the value of *a* and *b*. This value will be substitute into Equation (46) to obtain the least square linear fitting to the given data sets. The final fitting line has minimum error

.

2.4.2 Non-Linear Regression

This method can be used when we required to fit the data that clearly not present the linear characteristic. Exponential model is one of the example of nonlinear regression.

Given $(x_{1,}\mathrm{y}_1),(x_{2,}\mathrm{y}_2),(x_{3,}\mathrm{y}_3),.........,(x_{n,}\mathrm{y}_n),$ best fit $\mathrm{y}=ae^{bx}$ to the data. The variables a and b are the constants of the experimental model. The residual at each data point x_i is

$$E_i = \mathrm{y}_i - ae^{bx_i} \tag{53}$$

The sum of the square of the residuals is

$$S_r = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (\mathrm{y}_i - ae^{bx_i})^2 \tag{54}$$

To find the constants a and b of the exponential model, we minimize S_r by differentiating with respect to a and b and equating the resulting equations to zero.

$$\frac{\partial S_r}{\partial a} = \sum_{i=1}^{n} 2(\mathrm{y}_i - ae^{bx_i})(-e^{bx_i}) = 0 \tag{55}$$

$$\frac{\partial S_r}{\partial b} = \sum_{i=1}^{n} 2(\mathrm{y}_i - ae^{bx_i})(-ax_i e^{bx_i}) = 0 \tag{56}$$

or

$$-\sum_{i=1}^{n} \mathrm{y}_i e^{bx_i} + a\sum_{i=1}^{n} e^{2bx_i} = 0 \tag{57}$$

$$\sum_{i=1}^{n} \mathrm{y}_i x_i e^{bx_i} - a\sum_{i=1}^{n} x_i e^{2bx_i} = 0 \tag{58}$$

Equations (57) and (58) are non-linear in a and b and thus not in a closed form to be solved as was the case for linear regression. Generally, iterative methods such as Gauss-Newton iteration method, method of steepest descent, Marquardt's method or direct can be used to find values a and b .

However, in this case, from Equation (57), a can be written explicitly in term of b as

$$a = \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}} \tag{59}$$

Substituting equation (59) in (58) gives

$$\sum_{i=1}^{n} y_i x_i e^{bx_i} - \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}} \sum_{i=1}^{n} x_i e^{2bx_i} = 0 \tag{60}$$

The equation is still a nonlinear equation in b and can be solved best by numerical methods such as the bisection method or the secant method.

2.4.3 Gaussian

The Gaussian fitting is defined as follows:

$$\hat{y} = \sum_{i=1}^{N} a_i e^{\left[-\left(\frac{x-b_i}{c_i}\right)^2\right]} \tag{61}$$

Where N is a number of terms used and x is an original data. The real coefficient a_i, b_i and c_i need to be determined by using least square method.

For example if we used Gaussian fitting with one term and two terms, the Gaussian fitting equation is given by:

$$\hat{y}_1 = a_1 e^{\left[-\left(\frac{x-b_1}{c_1}\right)^2\right]} \tag{62}$$

and

$$\hat{y}_1 = a_1 e^{\left[-\left(\frac{x-b_1}{c_1}\right)^2\right]} + a_2 e^{\left[-\left(\frac{x-b_2}{c_2}\right)^2\right]} \tag{63}$$

2.4.4 Sine Fitting

$$\hat{y} = \sum_{j=1}^{M} a_j \sin(b_j x + c_j) \tag{64}$$

For example if we used Sine Fitting with one term and two terms, the equation is given by

$$\hat{y}_1 = a_1 \sin(b_1 x + c_1) \tag{65}$$

and

$$\hat{y}_2 = a_1 \sin(b_1 x + c_1) + a_2 \sin(b_2 x + c_2) \tag{66}$$

2.4.5 Fourier Series

$$\hat{y} = a_0 + \sum_{i=1}^{n} a_1 \cos(nwx) + b_1 \sin(nwx) \tag{67}$$

Similarly, if we used Fourier Fitting with one term and two terms, the equation is given by

$$\hat{y} = a_o + a_1 \cos(1wx) + b_1 \sin(1wx) \tag{68}$$

And

$$\hat{y} = a_o + a_1 \cos(1wx) + b_1 \sin(1wx) + a_2 \cos(2wx) + b_2 \sin(2wx) \tag{69}$$

2.4.6 Smoothing Spline

The smoothing spline s is constructed for the specified smoothing parameter,p and the specified weights, w_i . The smoothing spline minimizes

$$P\sum w_i(\mathrm{y}_i - s(x_i))^2 + (1-p)\int\left(\frac{d^2 s}{dx^2}\right)^2 dx \qquad (70)$$

They will be assumed as 1 for all data points if the weights are not specified. P can be defined between 0 and 1. Least-squares straight line will be constructed if $p = 0$ while $p = 1$ produces a cubic spline interpolation. Thus p value can be choosen between 0 and 1. The best value of p totally depends to the data and the experience of the user. Usually we need to change the value of p until we achieved the desired results

CHAPTER 3

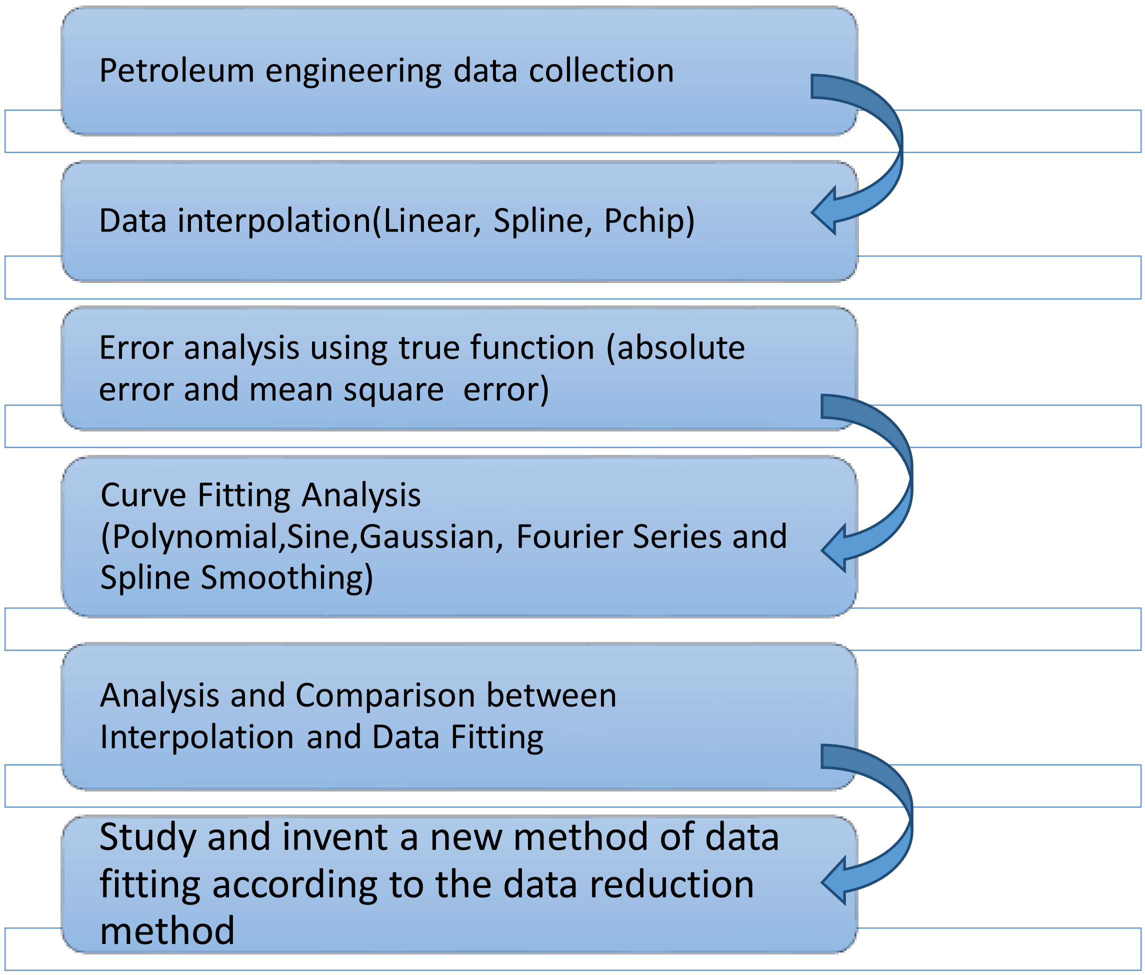
METHODOLOGY/PROJECT WORK

3.1 PETROLEUM ENGINEERING: DATA COLLECTION

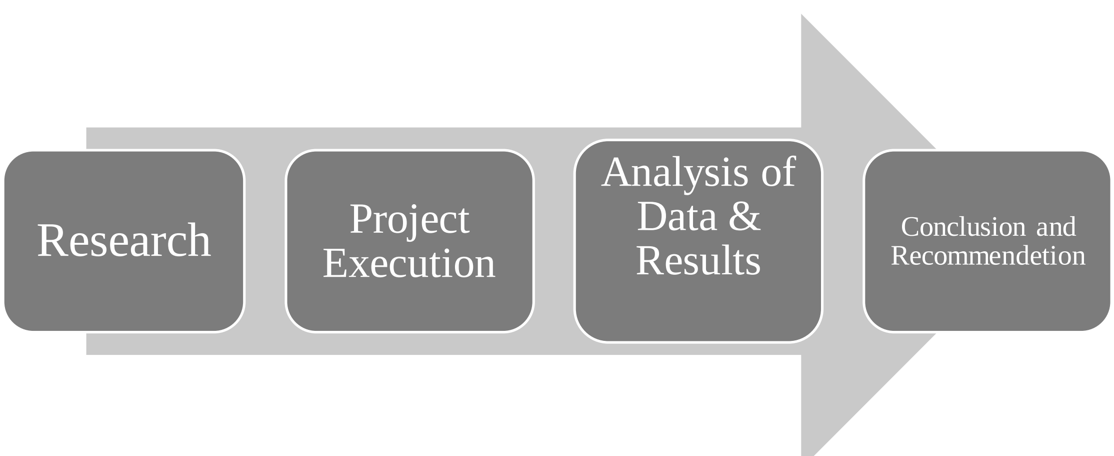
In petroleum engineering area either reservoir or drilling, every data collected will show in sparse distributed order which would be in non-uniform distribution. So, it is hard to interpret the real behaviour of the data and this will lead to the well planning problem. From the data visualization, large amount of disparate and potentially complex information can be quickly analysed by the users. Data interpretation is important in petroleum industry (Zhenzhen et al., 2012). Decision making process is really important because wrong decision will lead to many negative effects such as high operating cost and time wasted. Most of data visualization in petroleum industry only has been used within the reservoir area. Drilling data also needed to be visualize because estimation and approximation of the data is important in the drilling plan.

In this project report, data from final well reports that being own by one of the Oil and Gas Company in Malaysia has been collected. Data collected contain every parameter for their drilling operation such as Rate of Penetration, Torque and Drag, Pump pressure and Fluid Flow. All of the data are importance because its need to be accurate during the planning to aim for a successful drilling operation.

3.2 Research Methodology



3.2 Project Activities



3.3 Gantt Chart

Project work/Task	Week																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Title Selection																												
Preliminery Research						1																						
Detailed Study and PE Application researches								2																				
Collecting Data																												
Interpolation of Data																												
Error Analysis													3	4														
Curve Fitting Study																												
Curve Fitting Application in PE																					5							
Data Analysis																												
Comparison between Interpolation and Data Fitting																									6	7		
Project Wrapped Up and Conclusion																										8	9	

1.Submission of extended proposal	4. Submission of interim report	7. Submission of dissertation
2.Proposal Defence	5. Submission of progress report	8. Submission of technical paper
3. Submission of interim draft report	6.Submission of final draft report	9. Viva

CHAPTER 4

NEW CURVE FITTING METHOD

This chapter discuss the new curve fitting method that can be used to fit the given finite data sets. The new curve fitting methods has the capability to outperform the existing interpolation methods (Cubic Spline and PCHIP) and fitting methods (Polynomial, Gaussian, Fourier, Sine and Smoothing Spline).

Generally speaking, our new curve fitting method have the following special features:

1. It have combination between cubic spline and PCHIP
2. It consists of data reduction. For example, we may have data reduction for about 50% etc. By having data reduction, we may reduce the size of the storage for the data sets.
3. Our new curve fitting method have less error compare to the existing fitting scheme. The error is minimum and acceptable when we compare with cubic spline and PCHIP.
4. New curve fitting method is implemented by using new algorithm for curve fitting. This algorithm is totally new and will give the best results.

4.1 Algorithm

Input : Data point, n segment : n-1

Output : interpolating comes with several intervals

Step 1: Identify all the intervals.

(a) Linear interval
(b) Not linear interval

Step 2: If the linear interval $[x_1, x_{N1}]$

Use PCHIP interpolation (maintained the first and last point)

If the data is not linear, we divide into several new intervals

For $[i = N1;\ i\ <\ N1;\ i++]$

Example on how to calculate intervals number.

If N=50, Interval = 49

By following the ceiling and floor function, K-factor (Interval Reduction Facor) will be used to divide intervals

$$\left[\frac{n-1}{k}\right]$$

If k=8

Then, $\left[\frac{49}{8}\right]$= 6.125

If $\lfloor 6.5 \rfloor$ = 6

Else $\lceil 6.5 \rceil$ = 7

Step 3: After divided the intervals, we may have M segments $[m < n - 1]$

Then for each segment j

For $(j = 1, j < M\ , j++)$

We remove few points

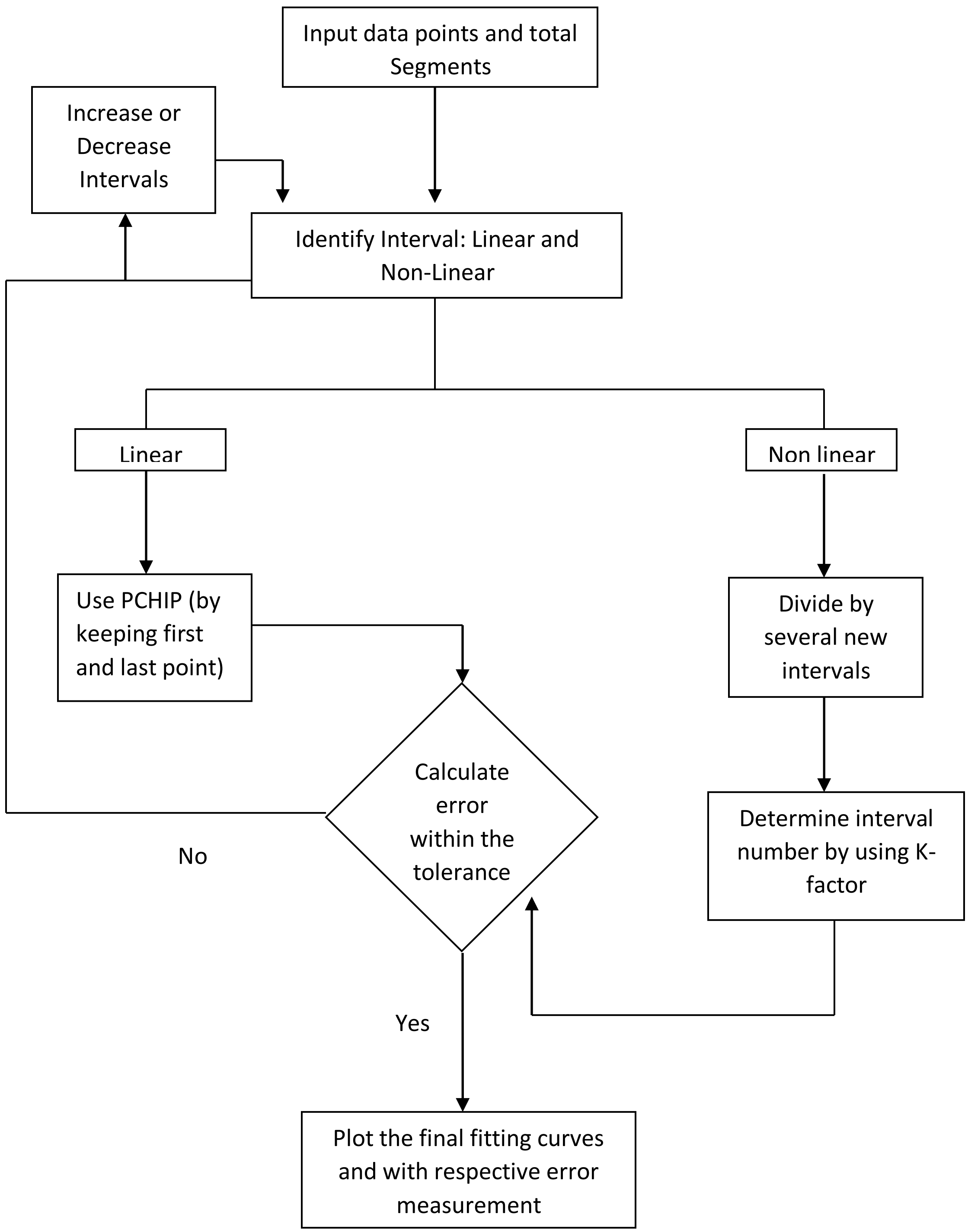
For $[i = n, i < m, i++]$

Step 4: Error calculation

Step 5: If the error statistic $(E = 1 \times 10^{-3})$, repeat **step 1**

The above algorithm can be represented by using the following schematic diagram in the next section.

4.2 Framework New Curve Fitting Method



CHAPTER 5

RESULTS AND DISCUSSION

5.1Data Interpolation

5.1.1 Measured Depth vs Torque

Torque measurement is important in drilling because too much torque will leads to the inability to reach the target.

1. Linear Interpolation

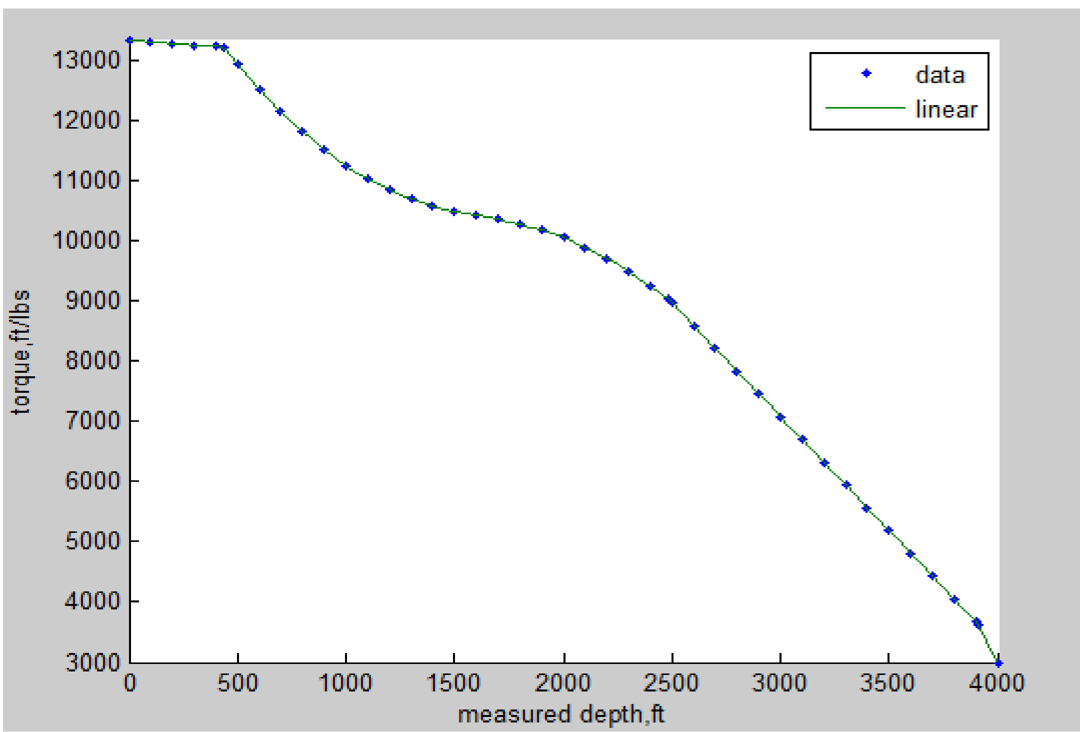


Figure 4: Linear Interpolation (Measured Depth vs Torque)

2. PCHIP Interpolation

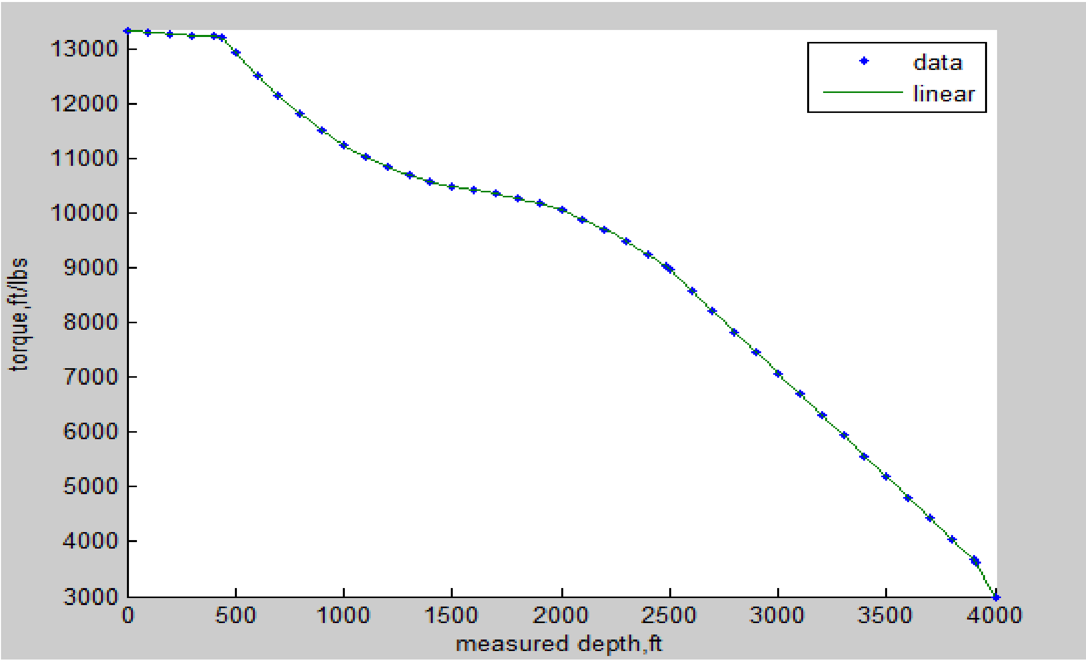


Figure 5: Pchip Interpolation (Measured Depth vs Torque)

3. Cubic Spline Interpolation

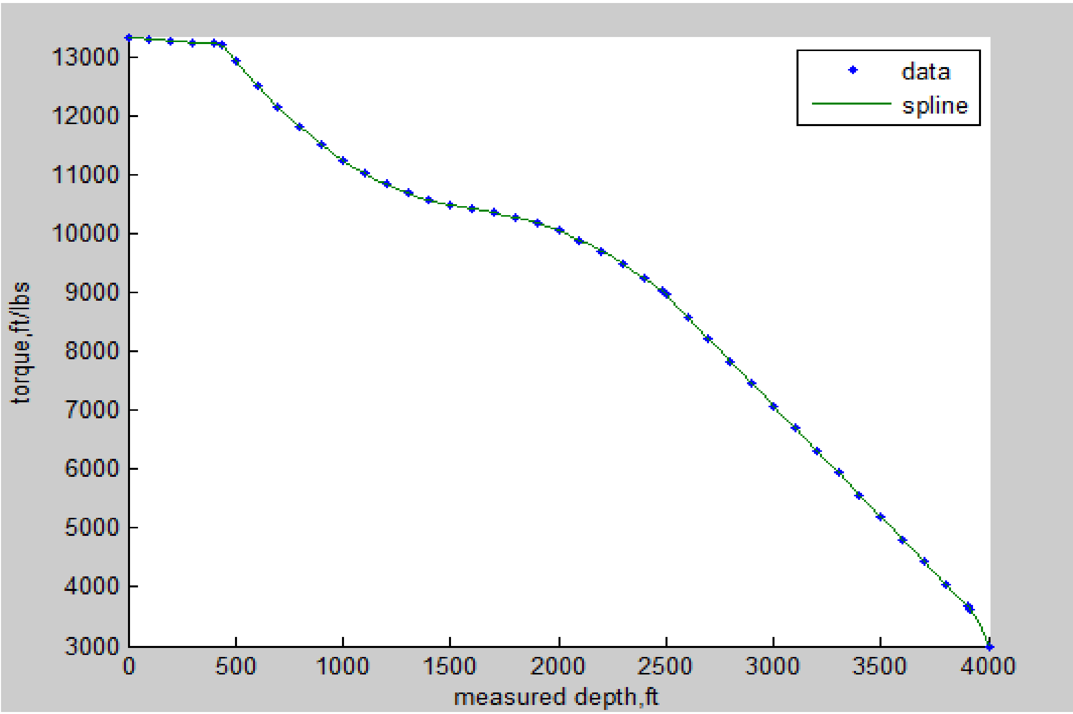


Figure 6: Cubic Spline Interpolation (Measured Depth vs Torque)

4. Combination of Linear, PCHIP and Spline Interpolation

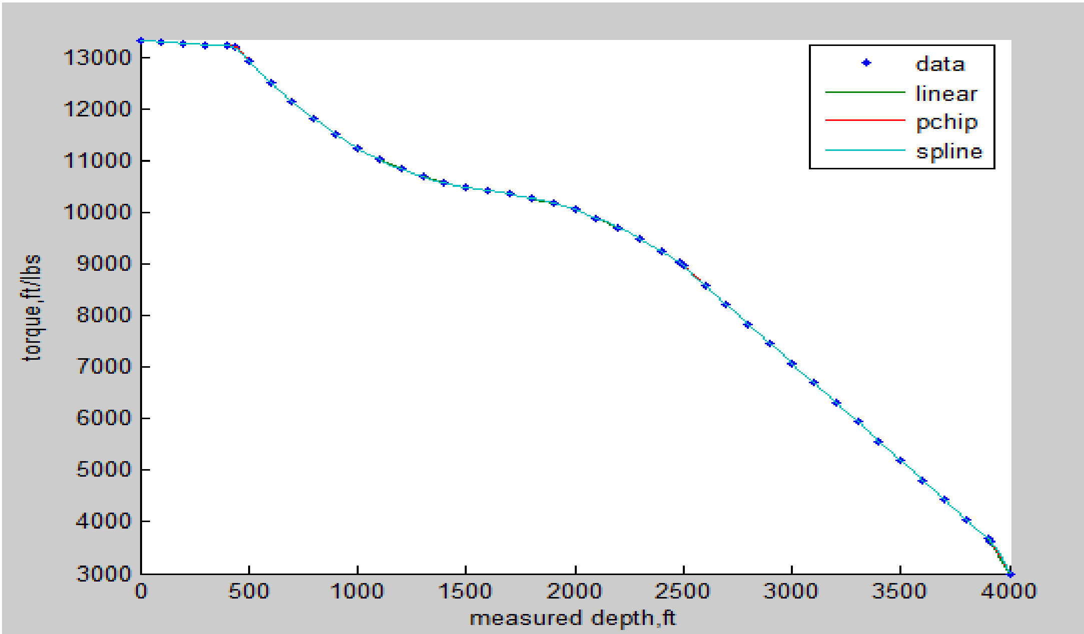


Figure 7: Combined Interpolation (Measured Depth vs Torque)

5.1.2 Measured Depth vs Fluid Flow

Fluid flow in the well is being measured to prevent any formation fluid flows into the well because if the if the fluid has been invaded, it will consider a kick and will lead to blow out of the well if not shut-in.

1. Linear Interpolation

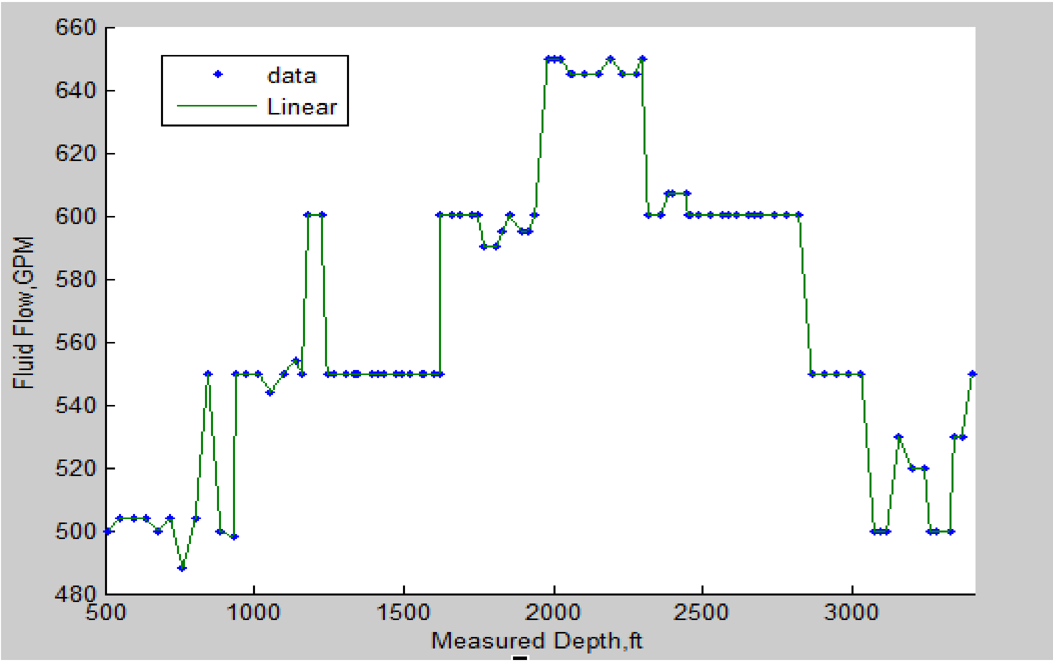


Figure 8: Linear Interpolation (Measured Depth vs Fluid Flow)

2. PCHIP Interpolation

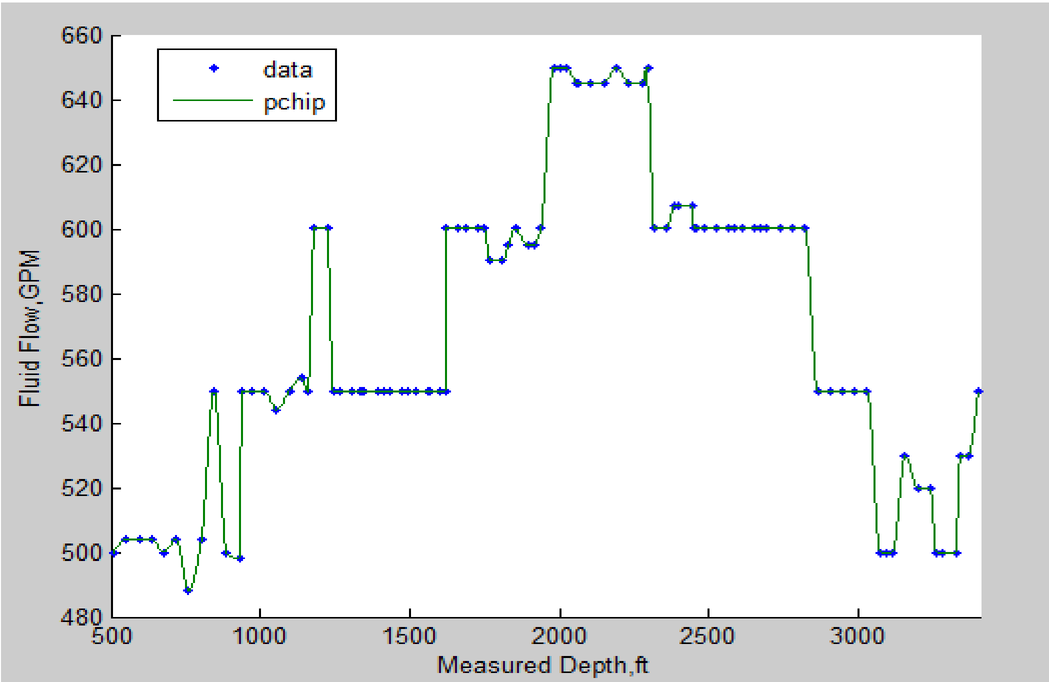


Figure 9: Pchip Interpolation (Measured Depth vs Fluid Flow)

3. Cubic Spline Interpolation

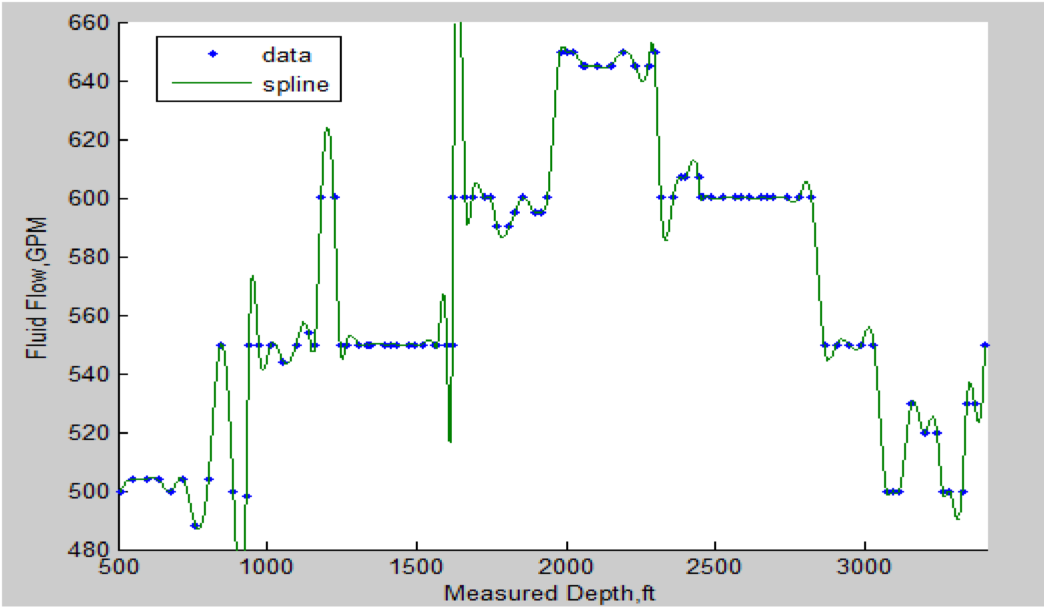


Figure 10: Cubic Spline Interpolation (Measured Depth vs Fluid Flow)

4. Combination of Linear, PCHIP and Spline Interpolation

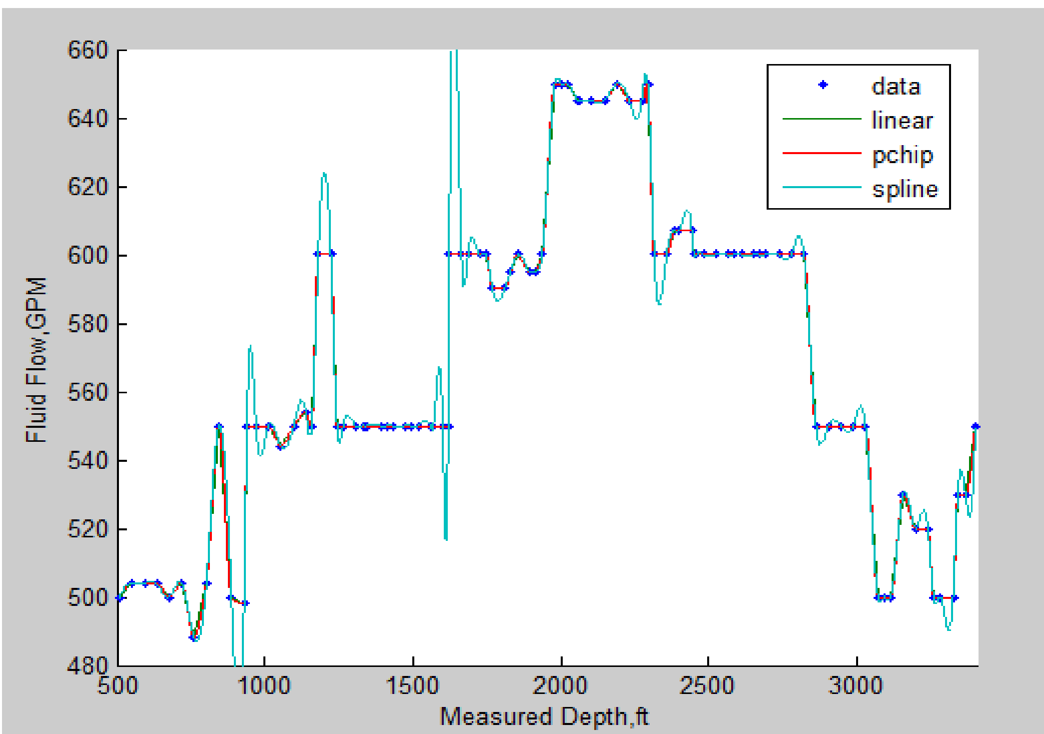


Figure 11: Combined Interpolation (Measured Depth vs Fluid Flow)

5.1.3 Data Interpolation Analysis

Table 3: Data Interpolation Analysis

Data Type	**Analysis**
Measured Depth vs Torque	Not much different can be seen through this interpolation because the torque data is distributed nicely and the interval between data points are small
Measured Depth vs Fluid Flow	The data have non-uniform distribution. So the differences between each spline interpolation can be seen. Cubic splines give the smoothest curve compared to pchip and linear interpolation.

5.2 Error Analysis

One of the techniques to measure the differences between the real values that have been observed with the values predicted by an estimator or a model is by using the root mean square error (RMSE). RMSE can be used to calibrate models to measure the drilling performance.

In this subsection, I use some examples of true function interpolation to calculate the absolute and root mean square error. The range of data 0 to 4 will be used to interpolate several functions. The formula to calculate the RMSE is show in Equation (71) below. The interpolation by using cubic spline and Pchip Interpolation will be compared in details.

$$\text{RMSE} = \sqrt{\frac{\sum_{n}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{71}$$

5.2.1 Function 1 ($f(x) = x^2 - 8$)

Table 4: Function 1 ($f(x) = x^2 - 8$)

x	$f(x)$	$S(x)$	$P(x)$	$S(x) - f(x)$	$(S(x) - f(x))^2$	$P(x) - f(x)$	$(P(x) - f(x))^2$
0.0	-8.00	-8.00	-8.0000	0	0	0.00000	0.000000
0.2	-7.96	-7.96	-7.9440	0	0	0.01600	0.000256
0.4	-7.84	-7.84	-7.7920	0	0	0.04800	0.002304
0.6	-7.64	-7.64	-7.5680	0	0	0.07200	0.005184
0.8	-7.36	-7.36	-7.2960	0	0	0.06400	0.004096
1.0	-7.00	-7.00	-7.0000	0	0	0.00000	0.000000
1.2	-6.56	-6.56	-6.6160	0	0	-0.05600	0.003136
1.4	-6.04	-6.04	-6.0880	0	0	-0.04800	0.002304
1.6	-5.44	-5.44	-5.4520	0	0	-0.01200	0.000144
1.8	-4.76	-4.76	-4.7440	0	0	0.01600	0.000256
2.0	-4.00	-4.00	-4.0000	0	0	0.00000	0.000000
2.2	-3.16	-3.16	-3.1867	0	0	-0.02670	0.000713
2.4	-2.24	-2.24	-2.2600	0	0	-0.02000	0.000400
2.6	-1.24	-1.24	-1.2400	0	0	0.00000	0.000000
2.8	-0.16	-0.16	-0.1467	0	0	0.01330	0.000177

3.0	1.00	1.00	1.0000	0	0	0.00000	0.000000
3.2	2.24	2.24	2.2187	0	0	-0.02130	0.000454
3.4	3.56	3.56	3.5360	0	0	-0.02400	0.000576
3.6	4.96	4.96	4.9440	0	0	-0.01600	0.000256
3.8	6.44	6.44	6.4347	0	0	-0.00530	0.000028
4.0	8.00	8.00	8.0000	0	0	0.00000	0.000000
			SUM	0	0	0.0000	0.02029

Absolute Error for Spline = 0

RMSE for spline = 0

Absolute Error for Pchip = 0

RMSE for pchip = $\sqrt{\dfrac{0.02029}{21}}$ = 0.031

1. Original Function Interpolation

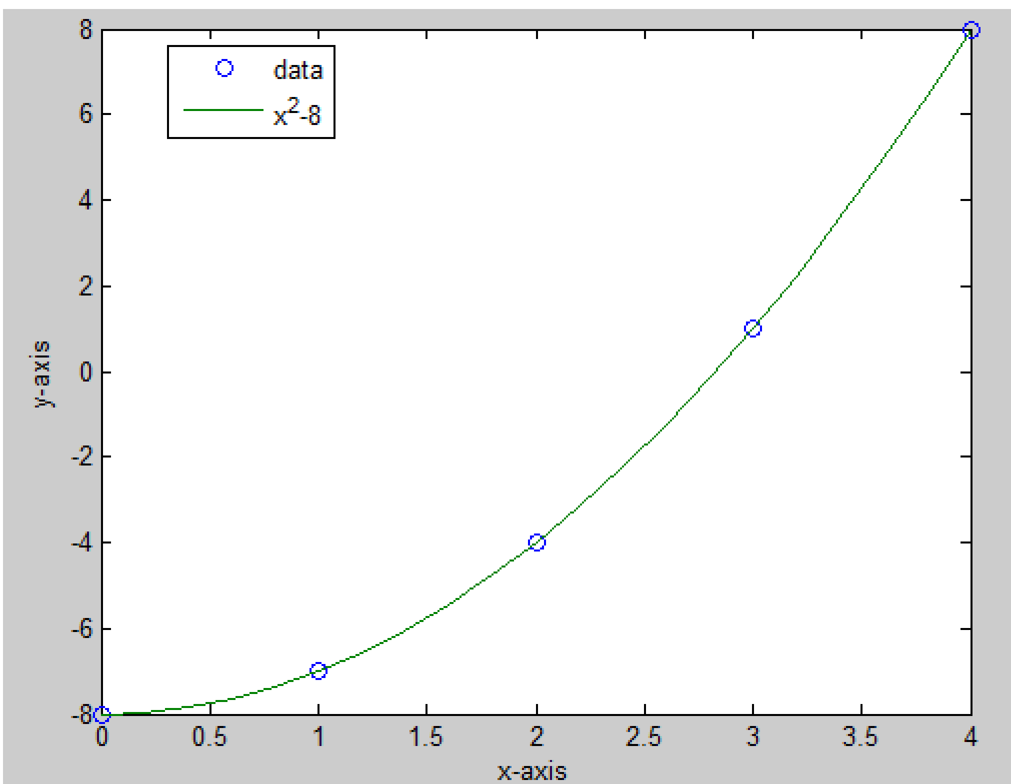


Figure 12 : Original function, ($f(x) = x^2 - 8$) in [0, 4].

2. PCHIP Interpolation

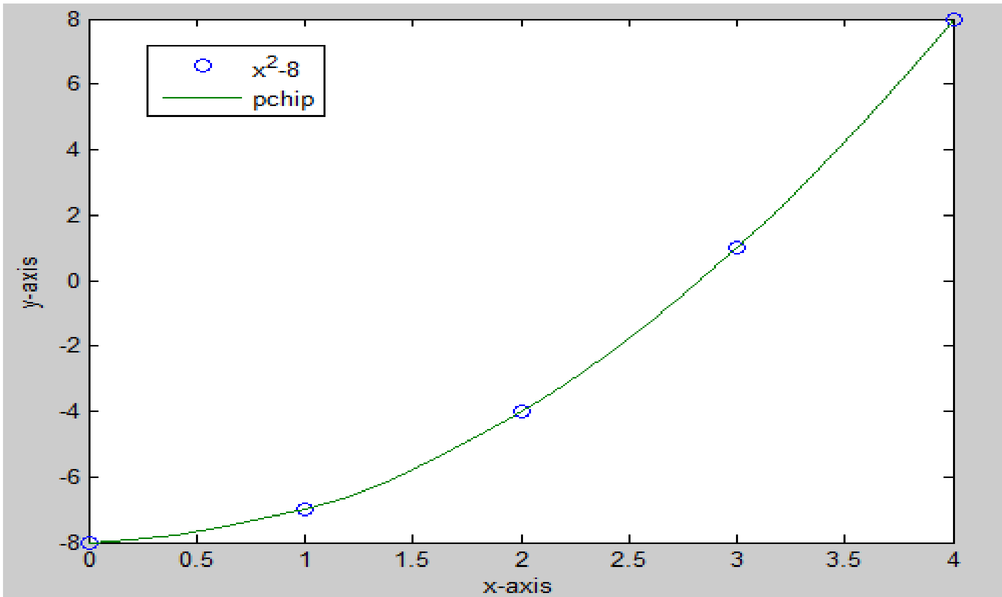


Figure 13: Pchip Interpolation, ($f(x) = x^2 - 8$) in [0,4].

3. Cubic Spline Interpolation

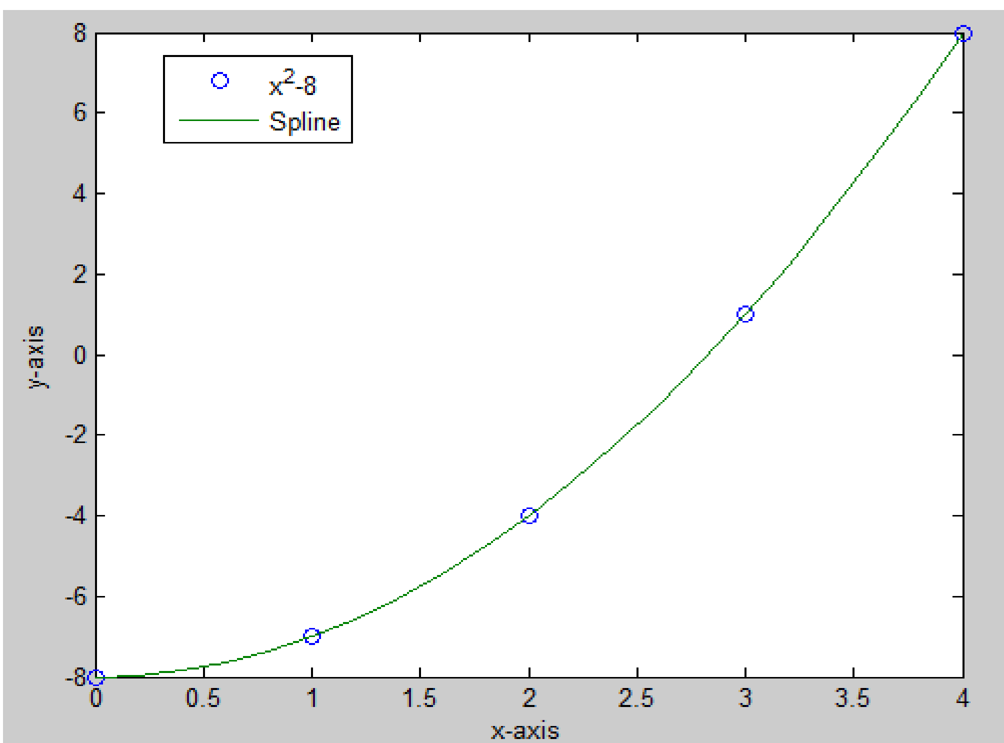


Figure 14: Spline Interpolation, ($f(x) = x^2 - 8$) in [0,4]

4. Combination of original function, PCHIP and spline interpolation

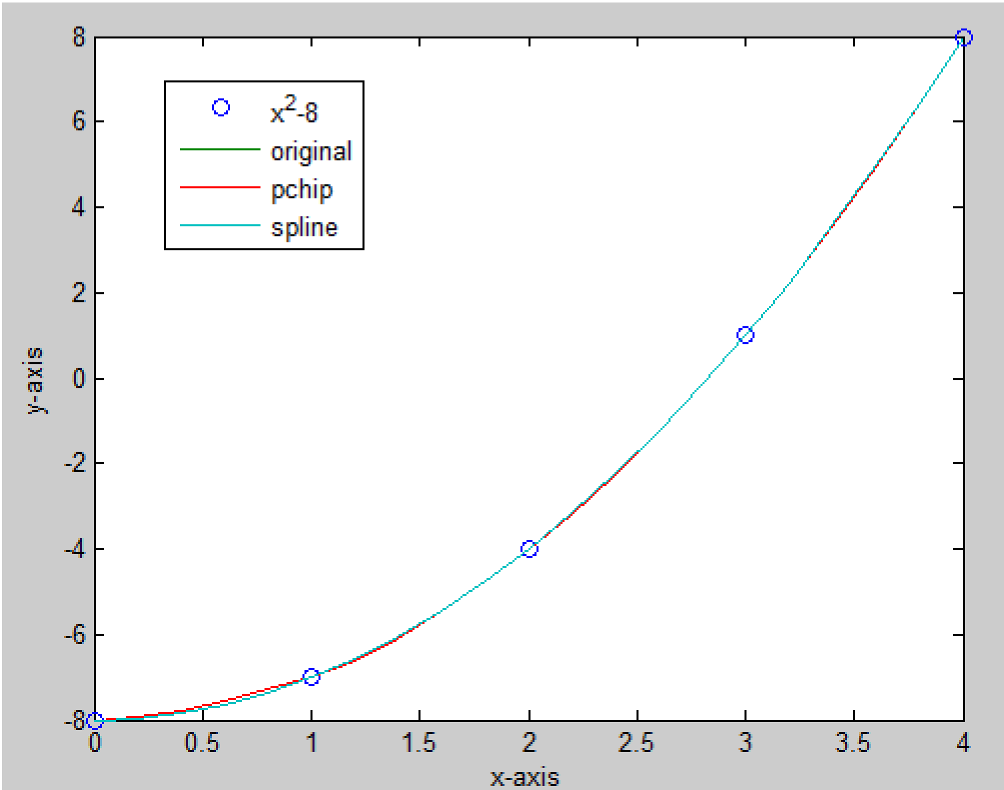


Figure 15: Combined Interpolation, ($f(x) = x^2 - 8$) in [0,4].

5.2.2 Function 2($f(x)=2e^x-x^2$)

Table 5: Function 2 ($f(x)=2e^x-x^2$)

x	$f(x)$	$S(x)$	$P(x)$	$S(x)-f(x)$	$(S(x)-f(x))^2$	$P(x)-f(x)$	$(P(x)-f(x))^2$
0.0	2.000	2.000	2.000	0.000000	0.000000	0.000000	0.000000
0.2	2.403	2.453	2.203	0.049895	0.002489	-0.200106	0.040042
0.4	2.824	2.876	2.589	0.052751	0.002783	-0.234249	0.054873
0.6	3.284	3.318	3.118	0.033262	0.001106	-0.165838	0.027502
0.8	3.811	3.822	3.748	0.011018	0.000121	-0.063082	0.003979
1.0	4.437	4.437	4.437	0.000000	0.000000	0.000036	0.000000
1.2	5.200	5.207	5.237	0.007066	0.000050	0.036866	0.001359
1.4	6.150	6.181	6.247	0.030100	0.000906	0.096600	0.009332
1.6	7.346	7.403	7.491	0.056435	0.003185	0.144635	0.020919
1.8	8.859	8.920	8.993	0.060305	0.003637	0.133605	0.017850
2.0	10.778	10.778	10.778	0.000000	0.000000	-0.000012	0.000000
2.2	13.210	13.059	13.155	-0.150827	0.022749	-0.054927	0.003017
2.4	16.286	15.984	16.403	-0.302753	0.091659	0.116547	0.013583
2.6	20.168	19.807	20.502	-0.360876	0.130232	0.334124	0.111639
2.8	25.049	24.784	25.431	-0.265294	0.070381	0.381607	0.145624
3.0	31.171	31.171	31.171	0.000026	0.000000	0.000026	0.000000
3.2	38.825	39.224	38.900	0.398440	0.158754	0.074540	0.005556
3.4	48.368	49.197	49.471	0.828500	0.686412	1.102900	1.216388
3.6	60.237	61.346	62.381	1.109731	1.231503	2.144431	4.598585
3.8	74.962	75.928	77.124	0.965231	0.931671	2.161831	4.673513
4.0	93.196	93.196	93.196	0.000000	0.000000	0.000000	0.000000
			SUM	2.523011	3.337638	6.009535	10.943762

Absolute error for spline =2.52

RMSE for spline = $\sqrt{\dfrac{3.337638}{21}}$ = 0.40

Absolute error for pchip = 6.01

RMSE for pchip = $\sqrt{\dfrac{10.943762}{21}}$ = 0.72

1. Original Function Interpolation

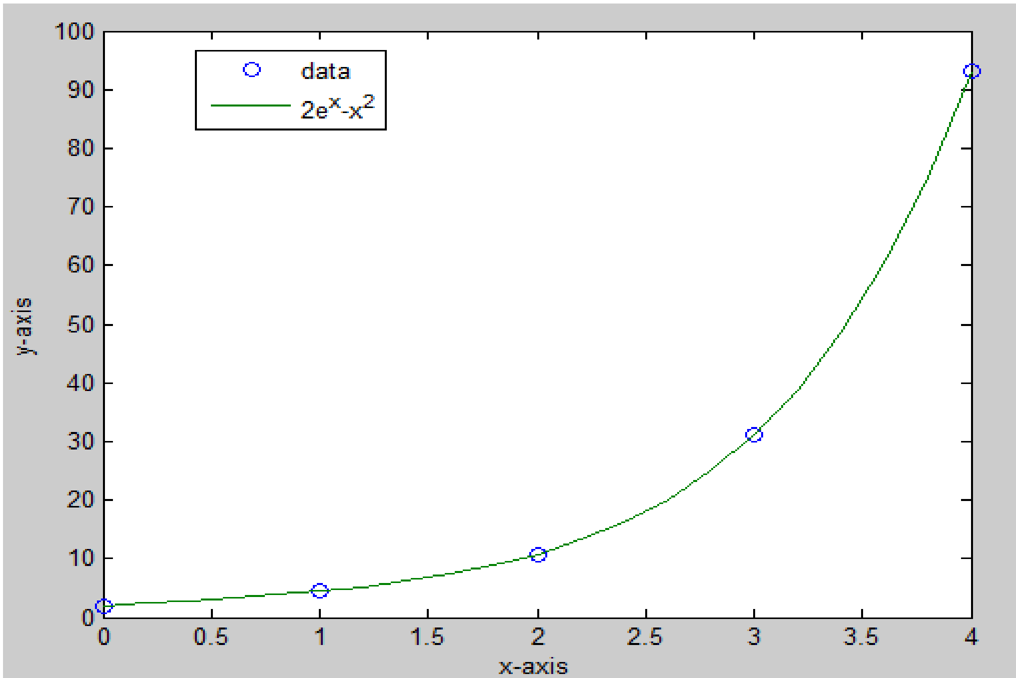


Figure 16: Original Function, ($f(x) = 2e^x - x^2$) in [0, 4].

2. PCHIP Interpolation

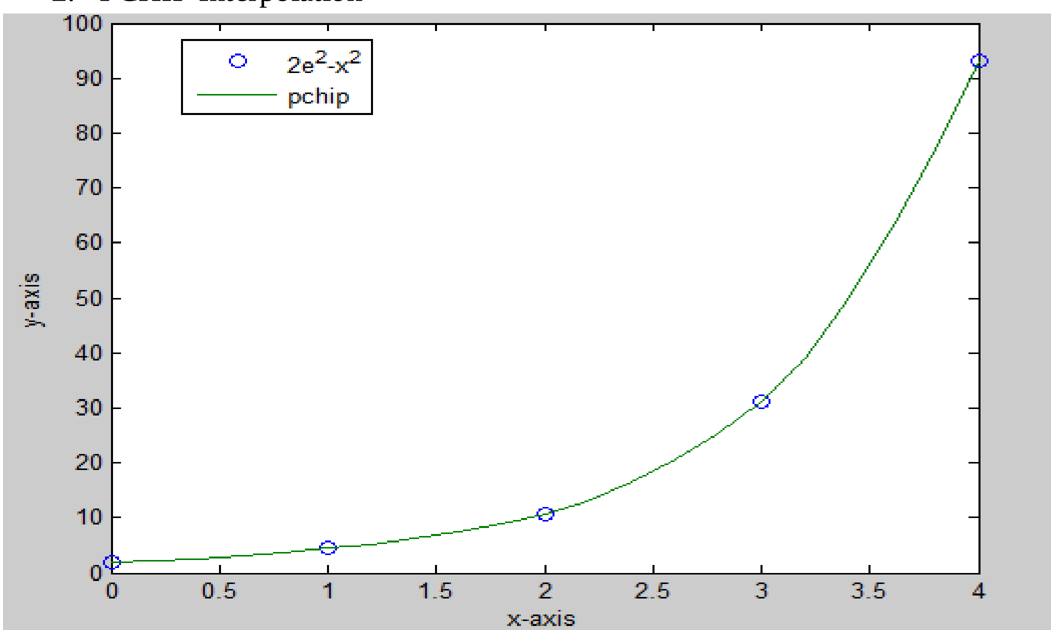


Figure 17: Pchip Interpolation, ($f(x) = 2e^x - x^2$) in [0, 4].

3. Cubic Spline Interpolation

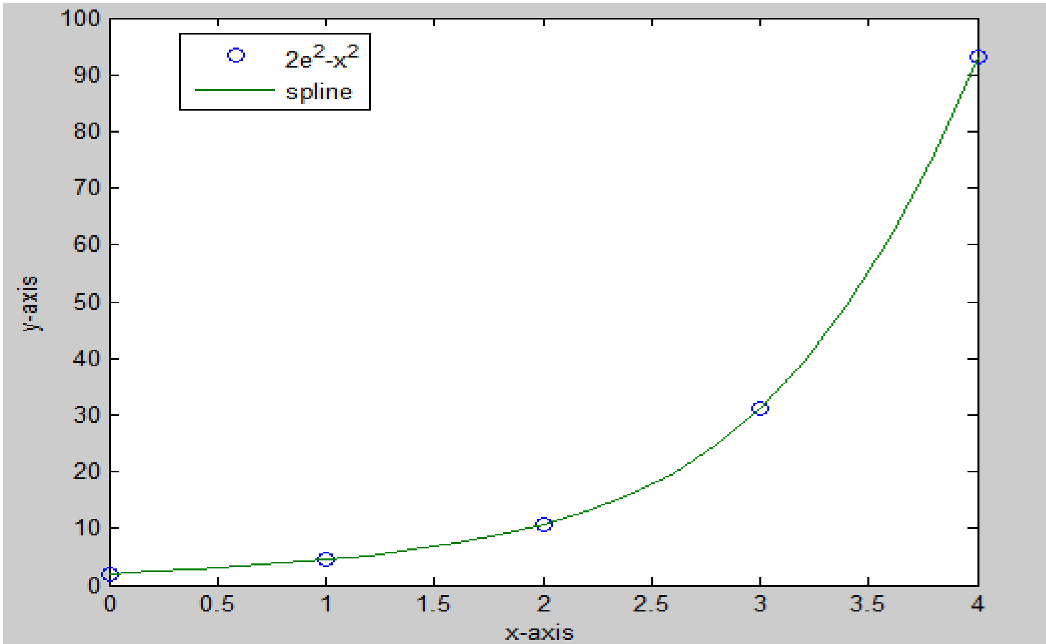


Figure 18: Spline Interpolation, ($f(x) = 2e^x - x^2$) in [0, 4].

4. Combination of original function, PCHIP and spline interpolation

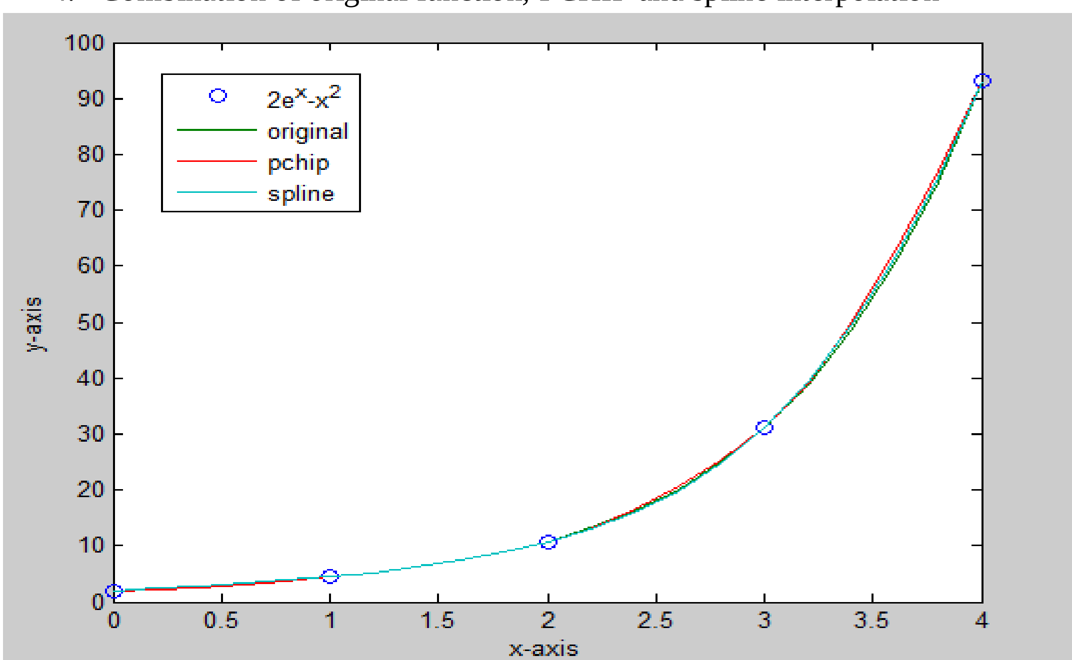


Figure 19: Combined Interpolation ($f(x) = 2e^x - x^2$) in [0, 4].

5.2.3 Function 3($f(x) = \cos^{10}(x)$)

Table 6: Function 3 ($f(x) = \cos^{10}(x)$)

x	$f(x)$	$S(x)$	P(x)	$S(x)-f(x)$	$(S(x)-f(x))^2$	$P(x)-f(x)$	$(P(x)-f(x))^2$
0.0	1.000	1.000	1.000	0.00000	0.00000	0.00000	0.00000
0.2	0.818	0.748	0.705	-0.06983	0.00488	-0.11273	0.01271
0.4	0.439	0.517	0.434	0.07798	0.00608	-0.00572	0.00003
0.6	0.147	0.313	0.210	0.16664	0.02777	0.06364	0.00405
0.8	0.027	0.140	0.059	0.11305	0.01278	0.03155	0.00100
1.0	0.002	0.002	0.002	0.00000	0.00000	-0.00002	0.00000
1.2	0.000	0.096	0.001	-0.09604	0.00922	0.00136	0.00000
1.4	0.000	0.150	0.001	-0.14970	0.02241	0.00090	0.00000
1.6	0.000	0.155	0.001	-0.15460	0.02390	0.00050	0.00000
1.8	0.000	0.106	0.000	-0.10610	0.01126	0.00020	0.00000
2.0	0.000	0.000	0.000	0.00000	0.00000	0.00004	0.00000
2.2	0.005	0.164	0.009	0.15862	0.02516	0.00444	0.00002
2.4	0.048	0.363	0.318	0.31497	0.09920	0.27087	0.07337
2.6	0.213	0.570	0.586	0.35678	0.12729	0.37258	0.13881
2.8	0.552	0.760	0.810	0.20821	0.04335	0.25881	0.06699
3.0	0.904	0.904	0.904	-0.00001	0.00000	-0.00001	0.00000
3.2	0.983	0.977	0.869	-0.00598	0.00004	-0.11418	0.01304
3.4	0.713	0.951	0.763	0.23796	0.05662	0.04916	0.00242
3.6	0.336	0.800	0.585	0.46388	0.21518	0.24858	0.06179
3.8	0.096	0.497	0.336	0.40085	0.16068	0.23975	0.05748
4.0	0.014	0.014	0.014	0.00000	0.00000	-0.00004	0.00000
			SUM	1.91668	0.84583	1.30968	0.43170

Absolute error for spline = 1.92

RMSE for spline = $\sqrt{\dfrac{0.84583}{21}}$ = 0.20

Absolute error for pchip = 1.31

RMSE for pchip = $\sqrt{\dfrac{0.43170}{21}}$ = 0.14

1. Original function interpolation

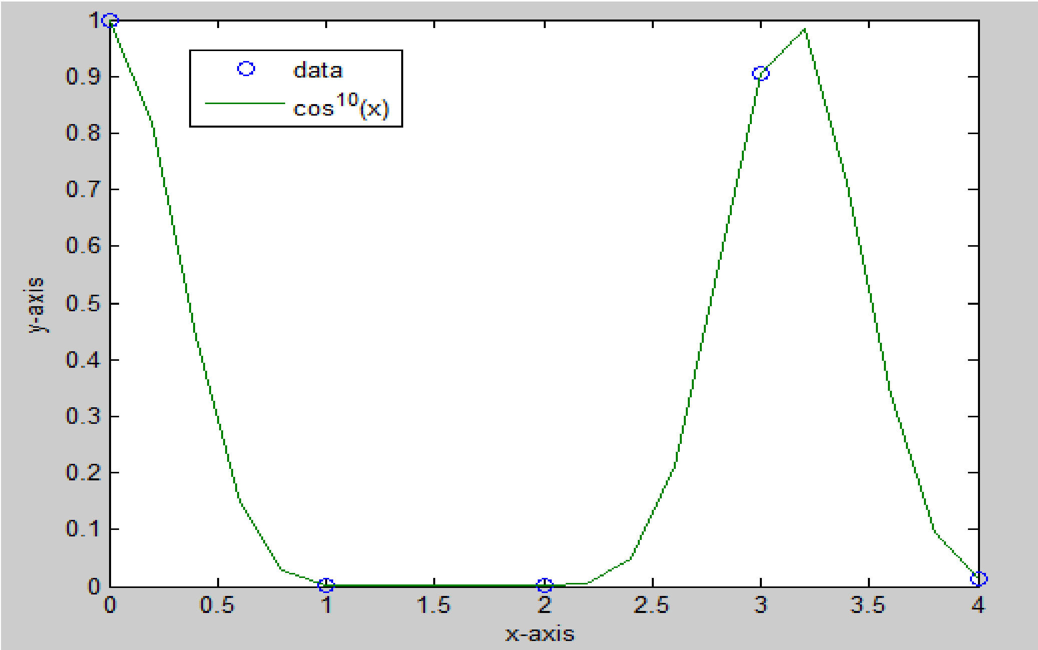


Figure 20: Original Function, ($f(x) = \cos^{10}(x)$) in [0,4].

2. PCHIP interpolation

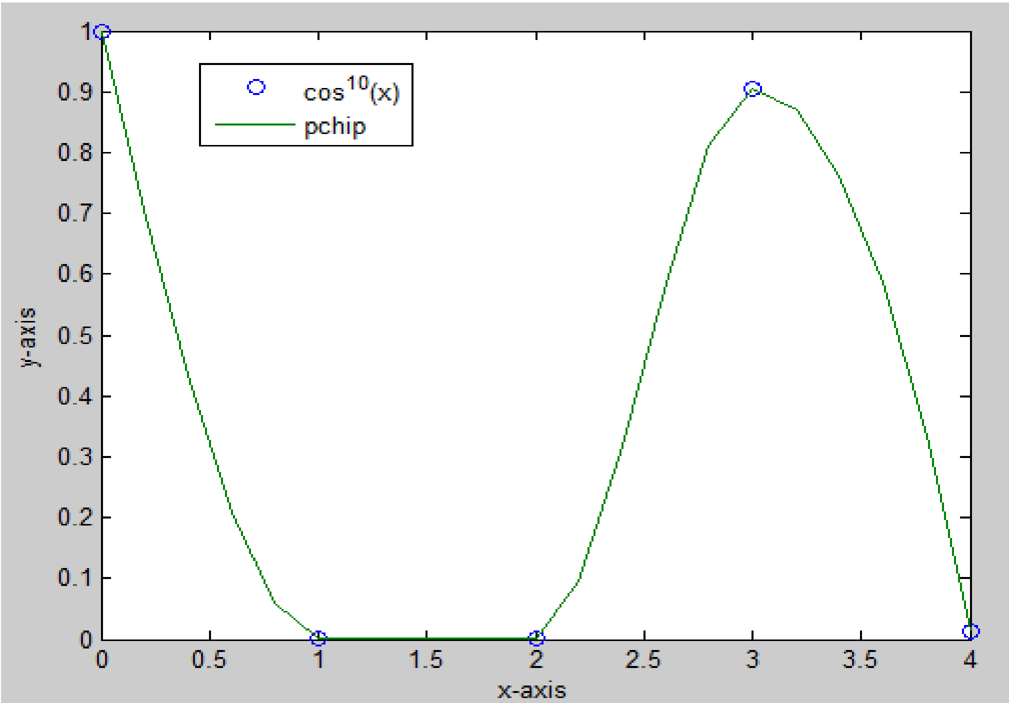


Figure 21: Pchip Interpolation ($f(x) = \cos^{10}(x)$) in [0,4].

3. Cubic Spline Interpolation

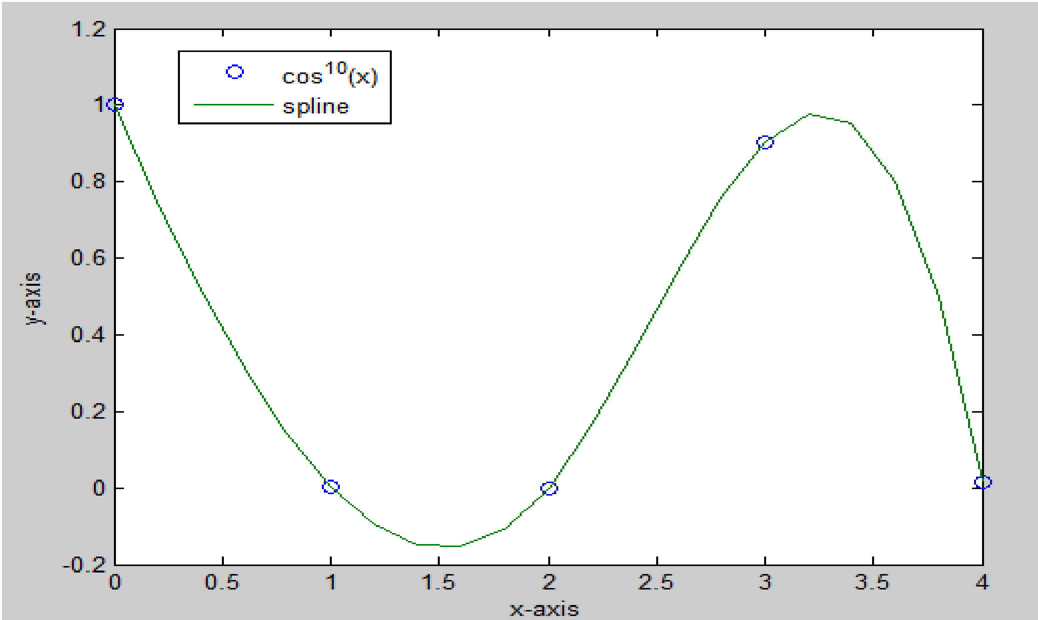


Figure 22: Spline Interpolation ($f(x) = \cos^{10}(x)$) in [0,4].

4. Combination of original function, PCHIP and spline interpolation

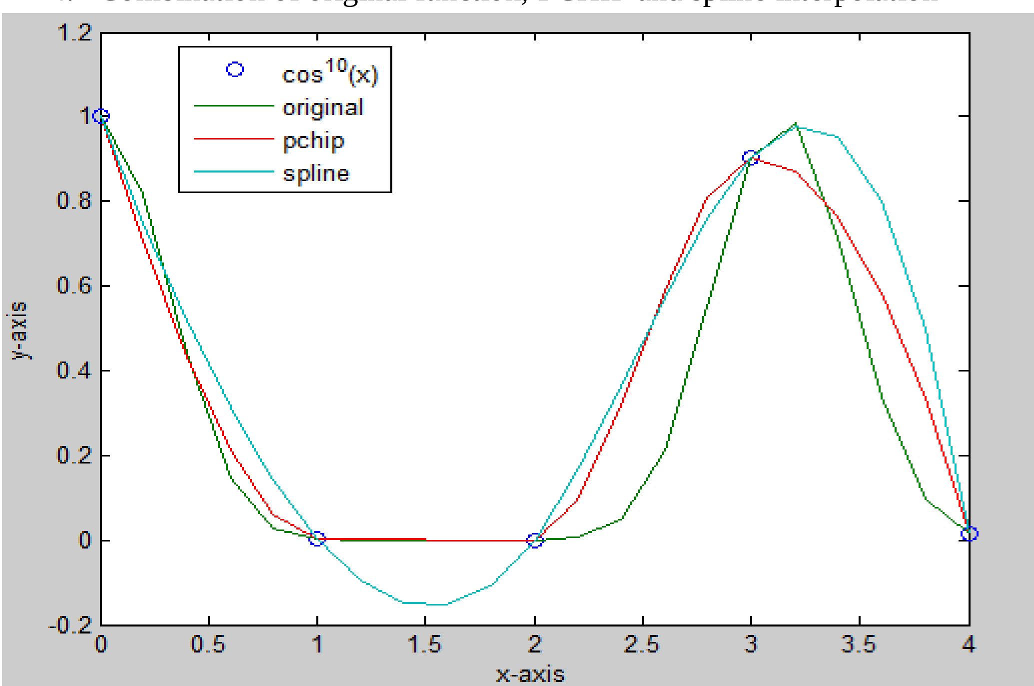


Figure 23: Combined Interpolation ($f(x) = \cos^{10}(x)$) in [0,4].

Table 7: Error analysis

Function	Absolute Error		Root Mean Square Error	
	PCHIP	Cubic Spline	PCHIP	Cubic Spline
$f(x) = x^2 - 8$	0	0	0.031	0
$f(x) = 2e^x - x^2$	6.01	2.52	0.72	0.40
$f(x) = \cos^{10}(x)$	1.31	1.92	0.14	0.20

From the Table 6, the absolute and root mean square error (RMSE) of cubic spline interpolations is lesser than the pchip interpolation for function $f(x) = x^2 - 8$ and $f(x) = 2e^x - x^2$ while for the function $f(x) = \cos^{10}(x)$, the absolute and RMSE of the cubic spline is higher that the pchip interpolation. Generally the cubic spline interpolation will give lesser error than the pchip interpolation

5.3 Curve Fitting Analysis

For this part, five(5) types of curve fitting method will be applied to the data which are Polynomial, Gaussian, Fourier, Sine and Smoothing Spline. Analysis of the best fit according to the number of error will be done to determine the best curve fitting method that is suitable for the used data sets.

5.3.1 Measured Depth vs Torque

Table 8: Measured Depth vs Torque (Curve Fitting)

Fitting Methods	Number Of Degree, n	Polynomial Equations	Root Mean Square Error (RMSE)
Polynomial	1	$f(x) = p_1 x + p_2$ p_1 = -2998 ,p_2 = 9226	619
	2	$f(x) = p_1 x^2 + p_2 x + p_3$ p_1 = -565.9, p_2= -3001 , p_3 = 9779	371.4
	3	$f(x) = p_1 x^3 + p_2 x^2 + p_3 x + p_4$ p_1 = 342.3 , p_2 = -557.6, p_3 = -2407, p_4 = 9770	273
	4	$f(x) = p_1 x^4 + p_2 x^3 + p_3 x^2 + p_4 x + p_5$ p_1 = 100.8, p_2 = -338.7, p_3 = -803.3, p_4 = -2415 , p_5 = 9839	268.7

	5	$f(x) = p_1x^5 + p_2x^4 + p_3x^3 + p_4x^2$ $+ p_5x + p_6$ p_1 = 367.4, p_2 = 128.4, p3 = -1493, p_4 = -861.2, p_5 = -1714, p_6 = 9841	181.6
	6	$f(x) = p_1x^6 + p_2x^5 + p_3x^4 + p_4x^3$ $+ p_5x^2 + p_6x + p_7$ p_1 = -364.4, p_2 = 323.3, p_3 = 1534, p_4 = -1373,p_5 = -2193, p_6 =-1770, p_7 =1.002e+04	74.49
	7	$f(x) = p_1x^7 + p_2x^6 + p_3x^5 + p_4x^4$ $+ p_5x^3 + p_6x^2 + p_7x + p_8$ p_1 = -100.1, p_2 = -376.1, p_3 = 775.6, p_4= 1570, p_5 = -1947 p_6= -2217 , p_7= -1594 p_8= 1.002e+04	70.75
	8	$f(x) = p_1x^8 + p_2x^7 + p_3x^6 + p_4x^5$ $+ p_5x^4 + p_6x^3 + p_7x^2 + p_8x + p_9$ p_1 =117.2 p_2 = -88.26 p_3 = -988.1 p_4 = 731.9 p_5 = 2559 p_6 = -1903 p_7 = -2722 p_8= -1603 p_9 = 1.006e+04	60.9
	9	$f(x) = p_1x^9 + p_2x^8 + p_3x^7 + p_4x^6$ $+ p_5x^5 + p_6x^4 + p_7x^3 + p_8x^2 + p_9x + p_{10}$ p_1 = -138.6 p_2=99.14 p_3 = 729.4 p_4 = -903.6 p_5 = -863.9 p_6 = 2437 p_7 = -763 p_8 = -2670 p_9= -1817 p_{10} = 1.006e+04	48.23

Gaussian	1	$f(x)=a_1e^{\left[-\left(\frac{x-b_1}{c_1}\right)^2\right]}$ a_1 = 1.296e+04 b_1 = -1.673 c_1 = 3.08	542.4
	2	$f(x)=a_1e^{\left[-\left(\frac{x-b_1}{c_1}\right)^2\right]}+a_2e^{\left[-\left(\frac{x-b_2}{c_2}\right)^2\right]}$ a_1 = 8152 b_1 = -1.708 c_1 = 0.7814 a_2 =1.011e+04 b_2 = -0.2173 c_2 =1.741	82.92
Fourier	1	$f(x)=a_o+a_1\cos(1wx)+b_1\sin(1wx)$ a_o = -2.337e+09 a_1 = 2.337e+09 b_1 = -4.312e+06 w = 0.0006959	376
	2	$f(x)=a_o+a_1\cos(1wx)+b_1\sin(1wx)+$ $a_2\cos(2wx)+b_2\sin(2wx)$ a_o = 8886 a_1 = 779.3 b_1 = -4420 a_2 = 213.9 b_2 = 1509 w = 1.14	172.1
Sine	1	$f(x)=a_1\sin(b_1x+c_1)$ a_1 = 1.342e+04 b_1 = 0.3362 c_1 = 2.327	402.3
	2	$f(x)=a_1\sin(b_1x+c_1)+a_2\sin(b_2x+c_2)$ a_1 = 2.935e+05 b_1 = 0.6467 c_1 =2.582 a_2 = 2.802e+05 b_2 = 0.6629 c_2 = -0.5482	280.4

Smoothing Spline	p = 0.5	314.7
	p = 0.6	286.2
	p = 0.7	258.2
	p = 0.8	228.3
	p = 0.85	211.3
	p = 0.9	190.8
	p = 0.95	159.3
	p = 0.99	0.03547

5.3.1.1Polynomial Curve Fitting Method

n=1

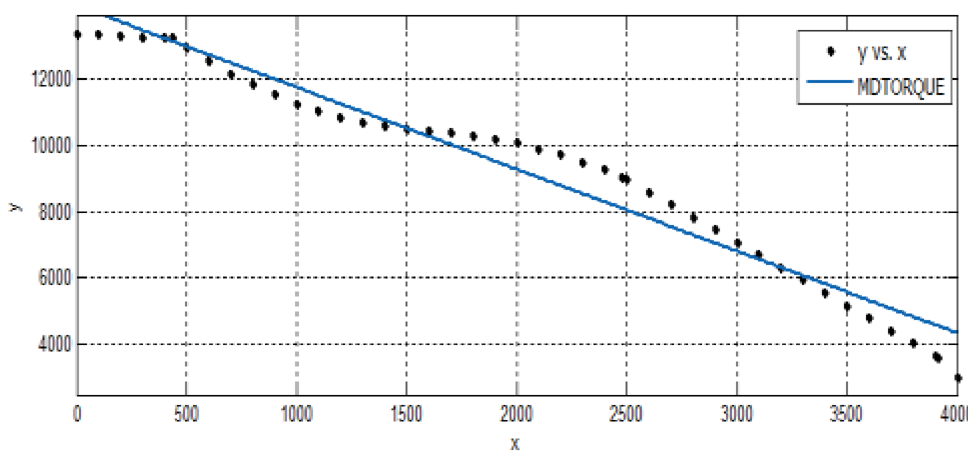


Figure 24: Polynomial Curve Fitting Method, n=1 (measured depth vs torque)

n=2

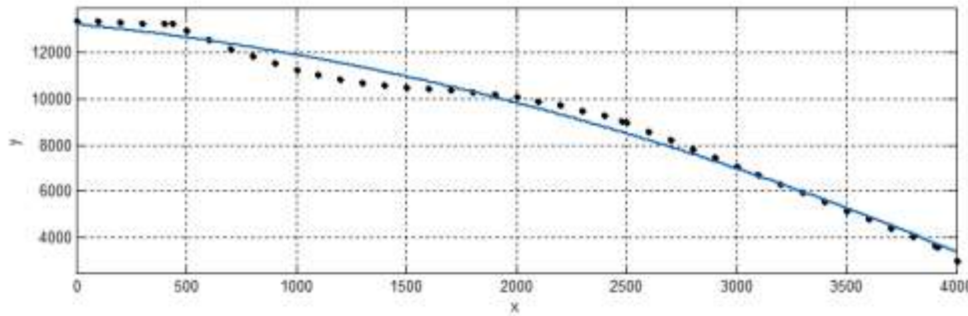


Figure 25: Polynomial Curve Fitting Method, n=2 (measured depth vs torque)

n=3

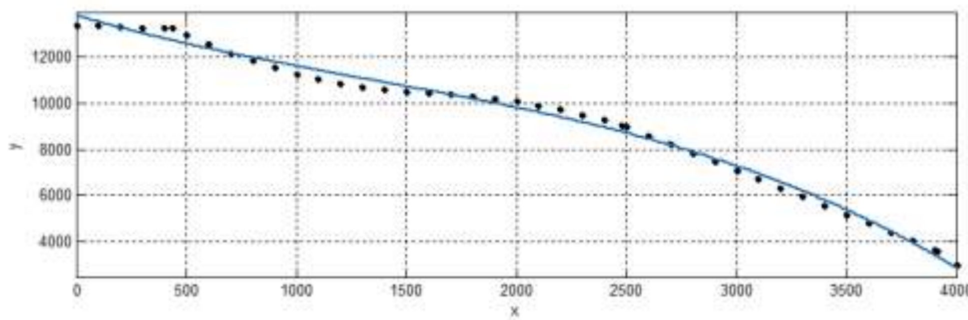


Figure 26: Polynomial Curve Fitting Method, n=3 (measured depth vs torque)

N=6

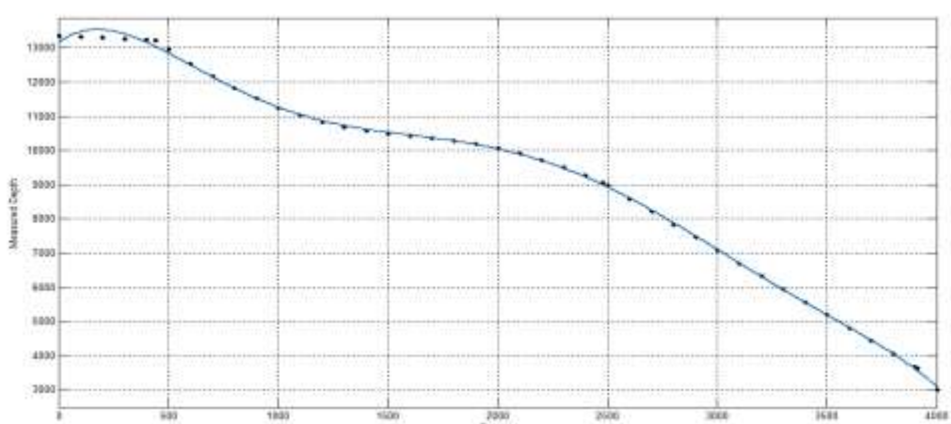


Figure 27: Polynomial Curve Fitting Method, n=6 (measured depth vs torque)

n=7

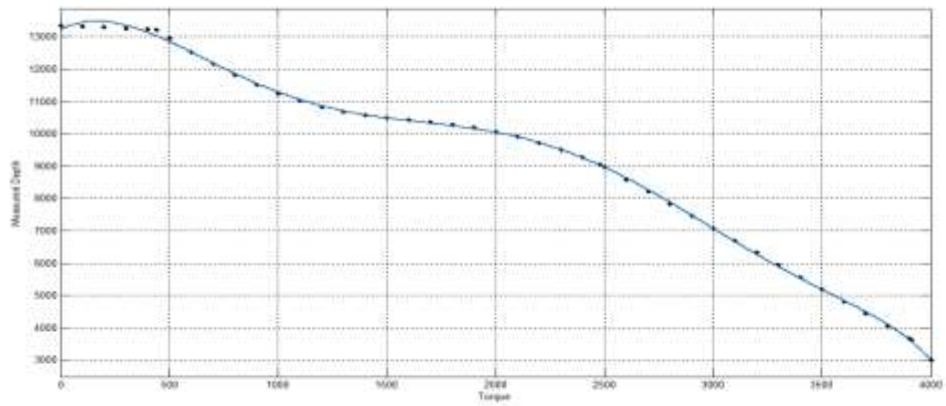


Figure 28: Polynomial Curve Fitting Method, n=7 (measured depth vs torque)

n=8

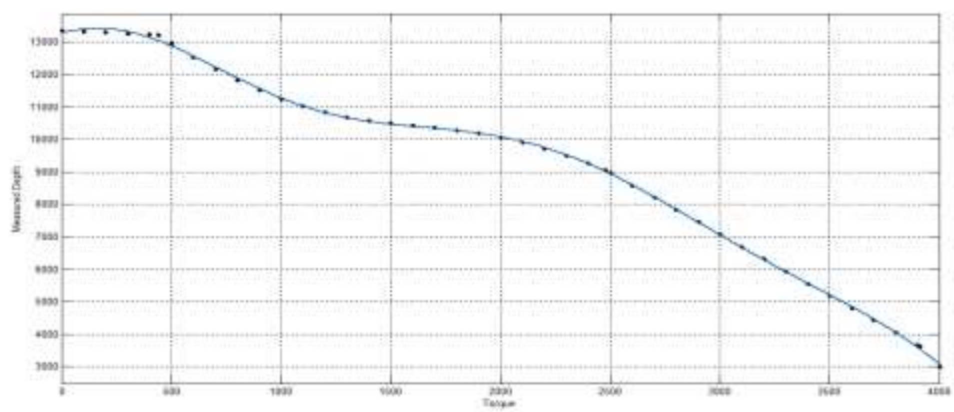


Figure 29: Polynomial Curve Fitting Method, n=8 (measured depth vs torque)

n=9

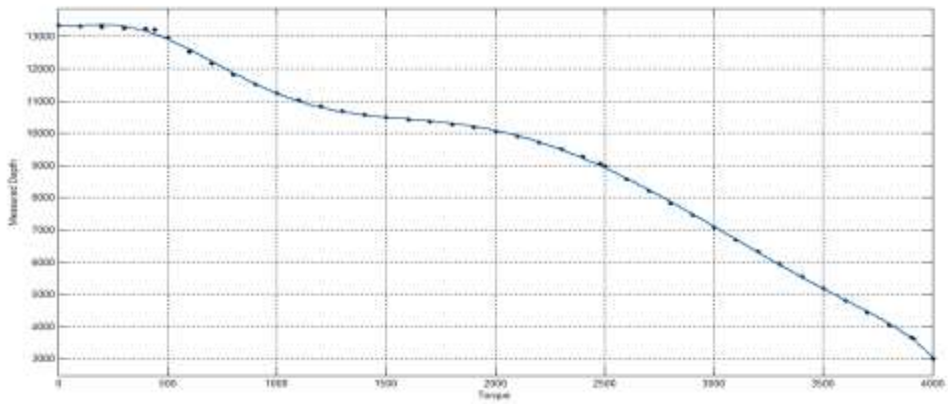


Figure 30: Polynomial Curve Fitting Method, n=9 (measured depth vs torque)

The results shown that the best for polynomial curve fitting is the degree 9. This shown that the best fit for this type of data is by using the higher degree number for polynomial curve fitting. In general, curve fitting using polynomial with higher degree is not recommended. This is due to the fact that the fitting curves may oscillate and it will destroy the characteristic of the data.

5.3.1.2 Gaussian Curve Fitting Method

n=1

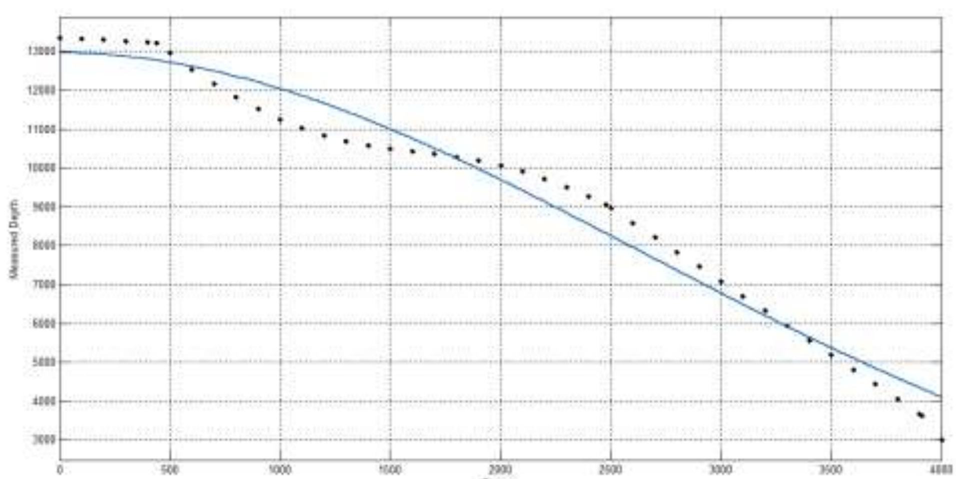


Figure 31: Gaussian Curve Fitting Method, n=1 (measured depth vs torque)

n=2

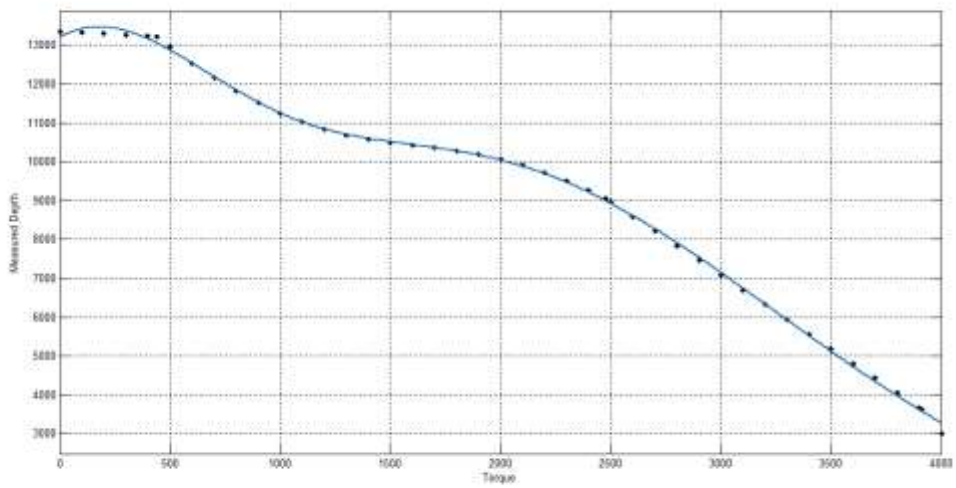


Figure 32: Gaussian Curve Fitting Method, n=2 (measured depth vs torque)

Gaussian method does not give the best fit for this type of data because it resulting in high error.

5.3.1.3 Fourier Curve Fitting Method

n=1

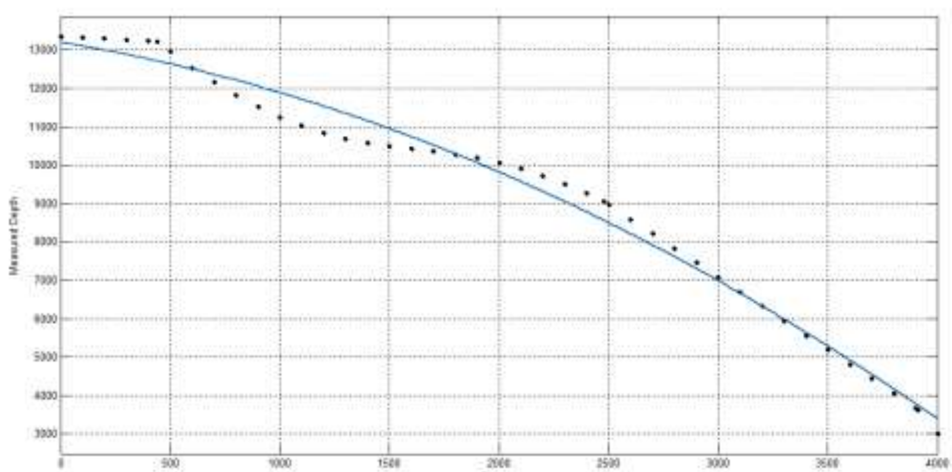


Figure 33: Fourier Curve Fitting Method, n=1 (measured depth vs torque)

n=2

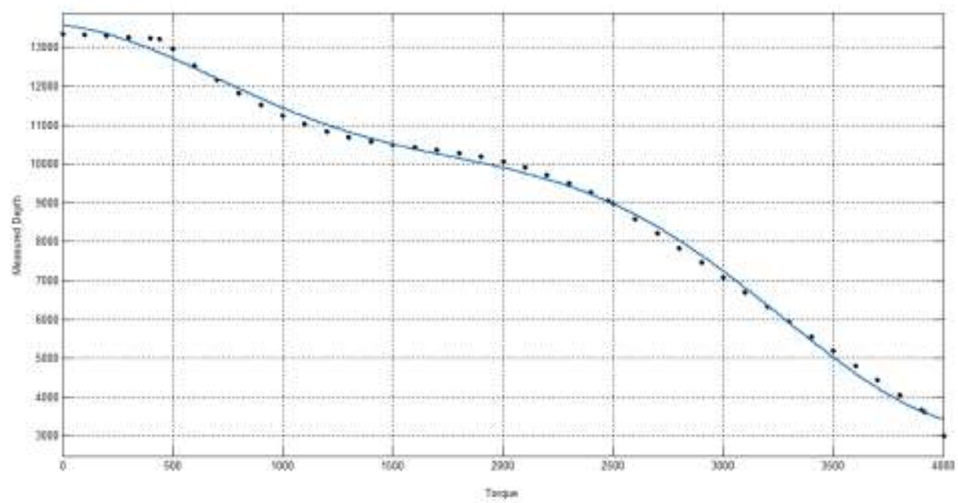


Figure 34: Fourier Curve Fitting Method, n=2(measured depth vs torque)

Fourier method does not give the best data because it resulting in a high error.

5.3.1.4 Sine Curve Fitting Method

n=1

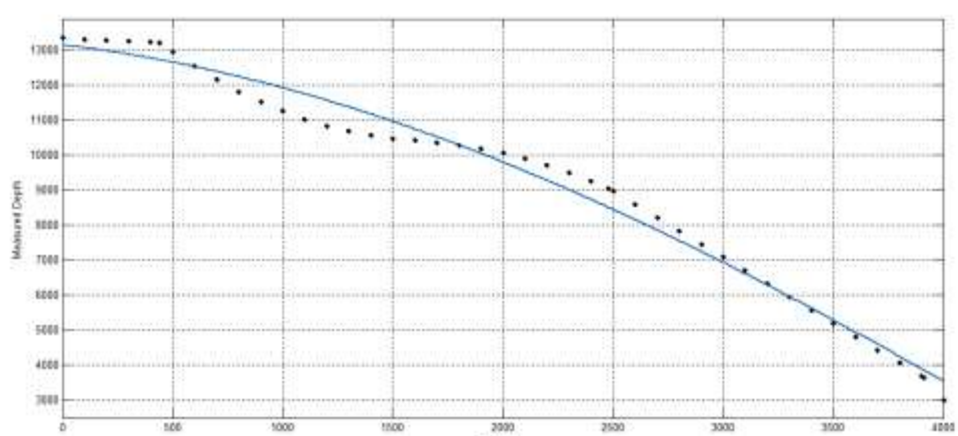


Figure 35: Sine Curve Fitting Method, n=1 (measured depth vs torque)

n=2

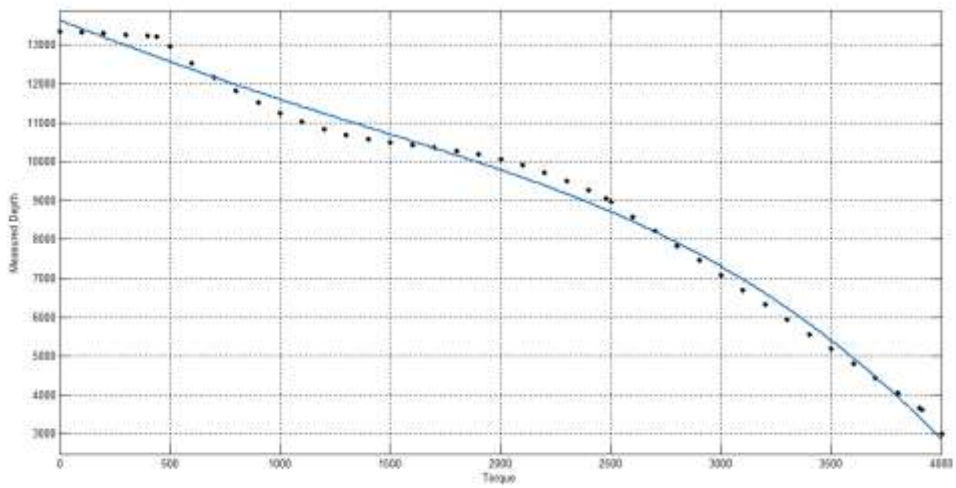


Figure 36: Sine Curve Fitting Method, n=2 (measured depth vs torque)

Sine does not give the best fit the data because this type of method resulting in high errors

5.3.1.5 Smoothing spline Curve Fitting Method

P=0.7

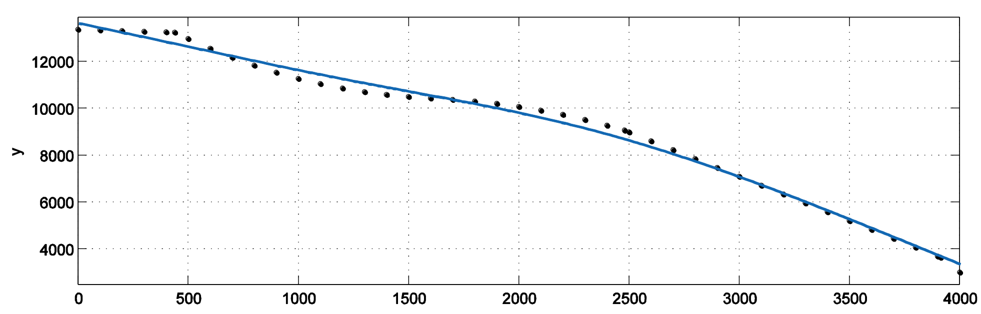


Figure 37: Smoothing Spline Curve Fitting Method, p=0.7 (measured depth vs torque)

P=0.8

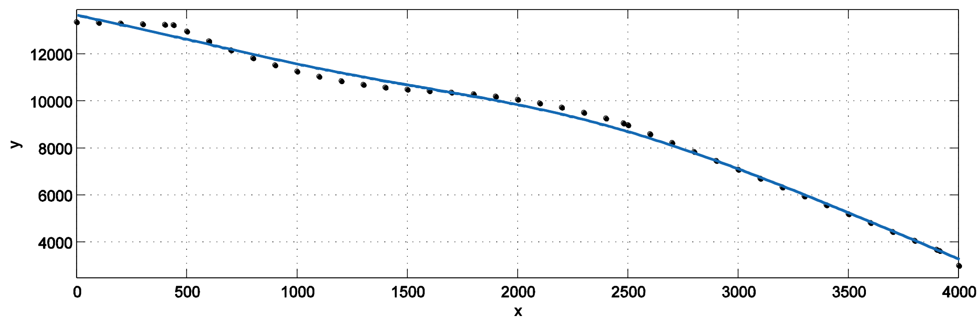


Figure 38: Smoothing Spline Curve Fitting Method, p=0.8 (measured depth vs torque)

P=0.9

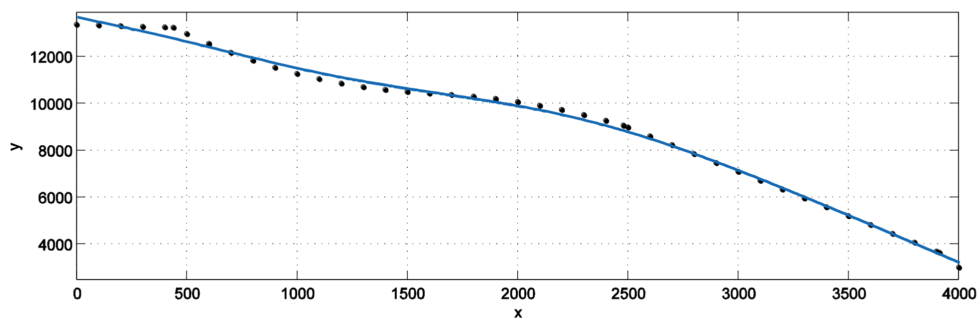


Figure 39: Smoothing Spline Curve Fitting Method, p=0.9 (measured depth vs torque)

P=0.95

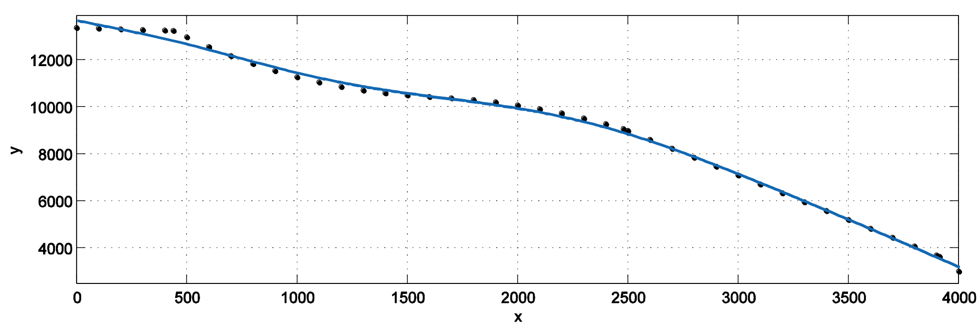


Figure 40: Smoothing Spline Curve Fitting Method, p=0.95(measured depth vs torque)

P=0.99

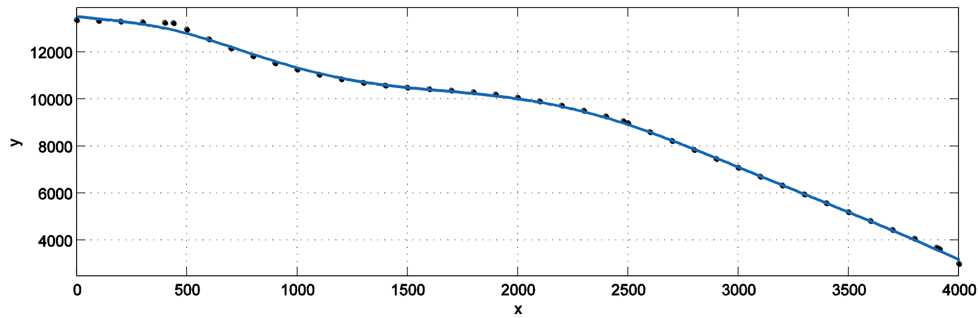


Figure 41: Smoothing Spline Curve Fitting Method, p=0.99 (measured depth vs torque)

For smoothing spline, the fit of the data is best when the value is closer to 1. Smoothing Spline gives the best fit and show the trend of the data without destroy the feature of the original data.

5.3.2 Measured Depth vs Fluid Flow

Table 9:Measured Depth vs Fluid Flow (curve fitting)

Fitting Methods	Number Of Degree, n	Polynomial Equations	Root Mean Square Error (RMSE)
Polynomial	1	$f(x) = p_1 x + p_2$ p_1 = -8.296 ,p_2 = 568.5	45.00
	2	$f(x) = p_1 x^2 + p_2 x + p_3$ p_1 = -38.46, p_2= 10.5, p_3 = 606.6	25.93
	3	$f(x) = p_1 x^3 + p_2 x^2 + p_3 x + p_4$ p_1 = -11.3, p_2 = -38.33, p_3 = 32.06 p_4 = 607.1	23.98
	4	$f(x) = p_1 x^4 + p_2 x^3 + p_3 x^2 + p_4 x$ $+ p_5$ p_1 = 10.67, p_2 = -11.01, p_3 = -67.94, p_4 = 31.31, p_5 = 616.2	22.55
	5	$f(x) = p_1 x^5 + p_2 x^4 + p_3 x^3 + p_4 x^2$ $+ p_5 x + p_6$ p_1 = 15.37, p_2 = 11.66 , p3 = -65.51 p_4 =-70.45, p_5 = 67.05 , p_6 = 617.2	19.61
		$f(x) = p_1 x^6 + p_2 x^5 + p_3 x^4 + p_4 x^3$ $+ p_5 x^2 + p_6 x + p_7$	19.57

	6	p_1 = 3.646, p_2 = 15.57 p_3 = -4.372 , p_4 =--66.06,p_5 = -53.36, p_6 =67.42, p_7 =614.6	
	7	$f(x) = p_1x^7 + p_2x^6 + p_3x^5 + p_4x^4 + p_5x^3 + p_6x^2 + p_7x + p_8$ p_1 = -5.942 , p_2 = 3.062, p_3 = 46.57, p_4= -2.012 p_5 = -111.2p_6= -56 , p_7= 83.09p_8= 615.1	19.35
	8	$f(x) = p_1x^8 + p_2x^7 + p_3x^6 + p_4x^5 + p_5x^4 + p_6x^3 + p_7x^2 + p_8x + p_9$ p_1 =14 p_2 = -4.5p_3 = -80.9 p_4 = 40.06 p_5 = 152.3 p_6 = -102.8 p_7 = -144p_8= 80.58 p_9 = 623	17.83
	9	$f(x) = p_1x^9 + p_2x^8 + p_3x^7 + p_4x^6 + p_5x^5 + p_6x^4 + p_7x^3 + p_8x^2 + p_9x + p_{10}$ p_1 = 7.364 p_2=14.9p_3 = -54.9 p_4 = -85.74 p_5 = 153.8 p_6 = 160.7 p_7 = -196.6p_8 = -148.9p_9= 101.3 p_{10} = 623.6	17.55
Gaussian	1	$f(x) = a_1 e^{\left[-\left(\frac{x-b_1}{c_1}\right)^2\right]}$ a_1 = 609.4 b_1 = 0.147 c_1 = 3.75	25.32
	2	$f(x) = a_1 e^{\left[-\left(\frac{x-b_1}{c_1}\right)^2\right]} + a_2 e^{\left[-\left(\frac{x-b_2}{c_2}\right)^2\right]}$ a_1 = 619.4 b_1 = 0.3292 c_1 = 2.905 a_2 =123.2 b_2= -1.851 c_2 =0.8094	22.65

Fourier	1	$f(x) = a_o + a_1\cos(1wx) + b_1\sin(1wx)$ a_o = 566.6 a_1 = 52.91 b_1 = 21.41 w = 1.79	21.94
	2	$f(x) = a_o + a_1\cos(1wx) + b_1\sin(1wx) +$ $a_2\cos(2wx) + b_2\sin(2wx)$ a_o = 567.9 a_1 = 52.44 b_1 = 20.08 a_2 = -3.326 b_2 = 13.77 w = 1.824	19.76
Sine	1	$f(x) = a_1\sin(b_1x + c_1)$ a_1 = 607.9 b_1 = 0.362 c_1 = 1.52	25.74
	2	$f(x) = a_1\sin(b_1x + c_1) + a_2\sin(b_2x + c_2)$ a_1 = 3114 b_1 = 0.4317 c_1 =2.702 a_2 = 675 b_2 = 0.4839 c_2 = -0.2722	24.13
Smoothing Spline		p = 0.85	19.89
		p = 0.9	19.28
		p = 0.95	18.3
		p = 0.99	16.04

5.3.2.1 Polynomial Curve Fitting Method

n=1

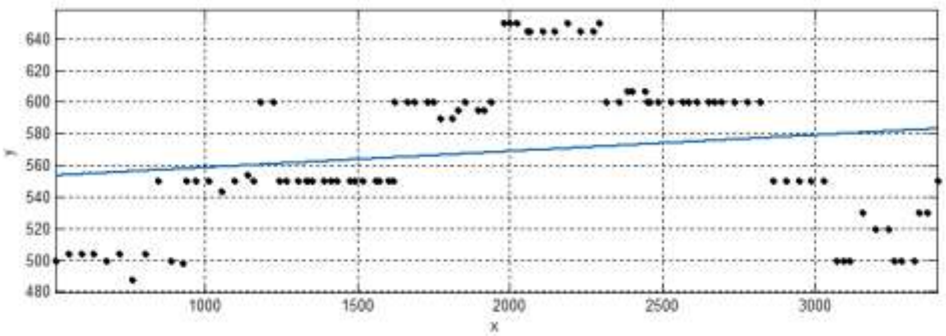


Figure 42: Polynomial Curve Fitting Method, n=1 (measured depth vs fluid flow)

n=2

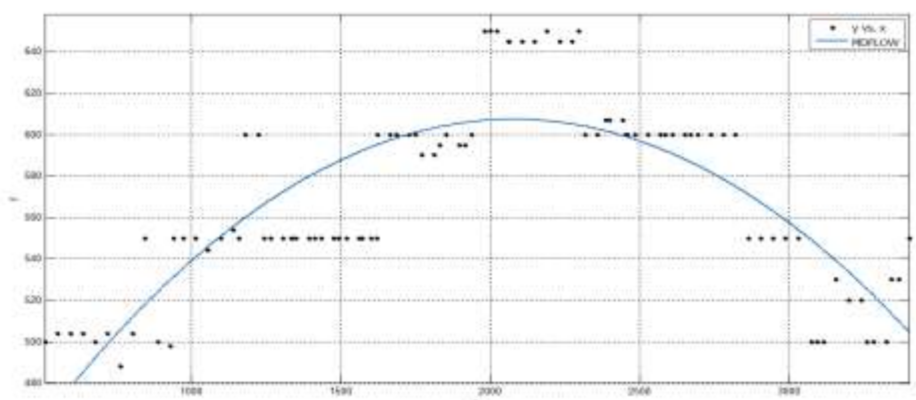


Figure 43: Polynomial Curve Fitting Method, n=2 (measured depth vs fluid flow)

n=3

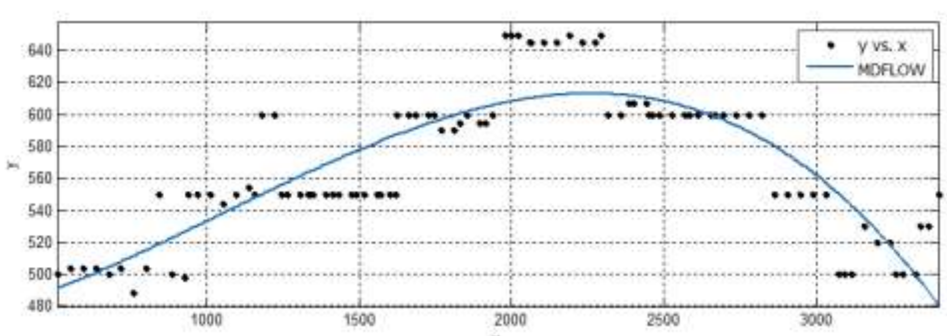


Figure 44: Polynomial Curve Fitting Method, n=3 (measured depth vs fluid flow)

n=4

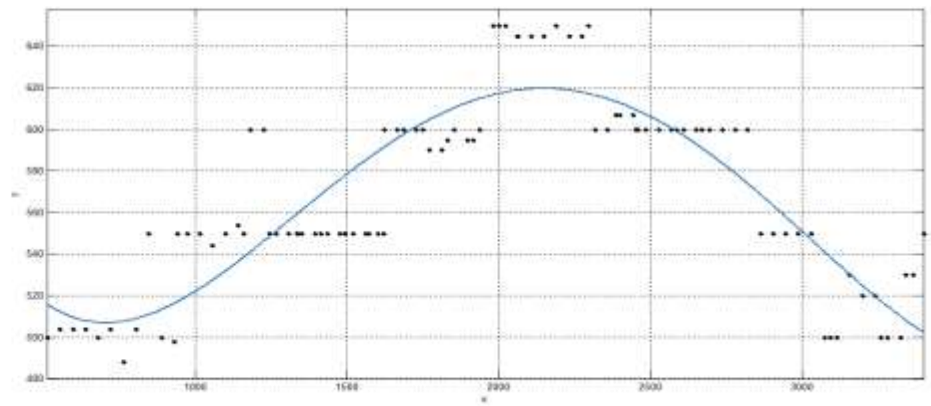


Figure 45: Polynomial Curve Fitting Method, n=4 (measured depth vs fluid flow)

n=7

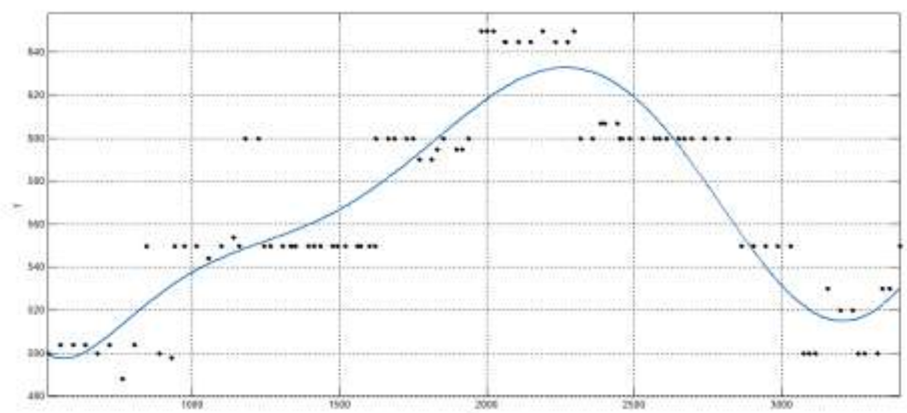


Figure 46: Polynomial Curve Fitting Method, n=7 (measured depth vs fluid flow)

n=8

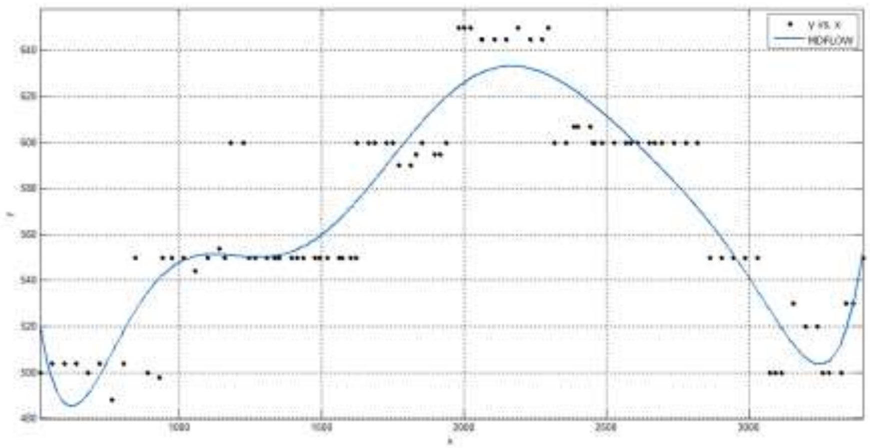


Figure 47: Polynomial Curve Fitting Method, n=8 (measured depth vs fluid flow)

n=9

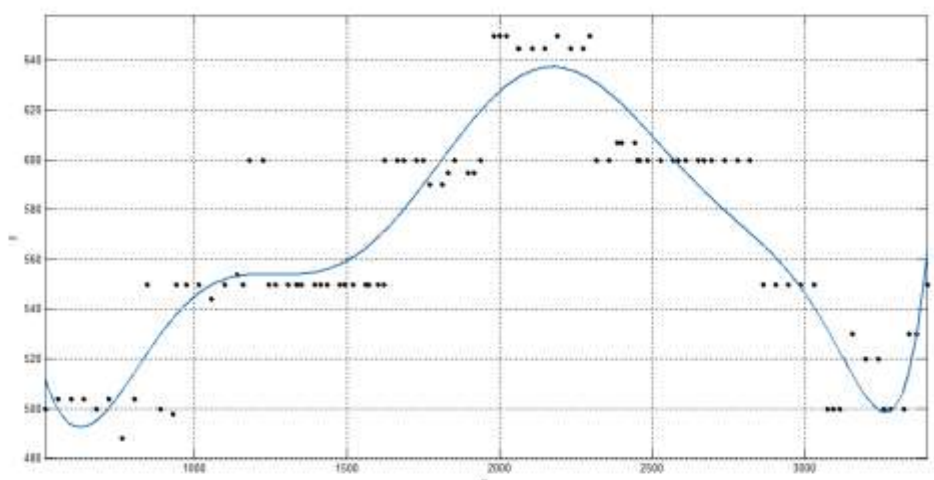


Figure 48: Polynomial Curve Fitting Method, n=9 (measured depth vs fluid flow)

For this data, polynomial fitting method is not suitable for all number of degree because the error too large and the fitting curves deviated from the original data.

5.3.2.2 Gaussian Curve Fitting Method

n=1

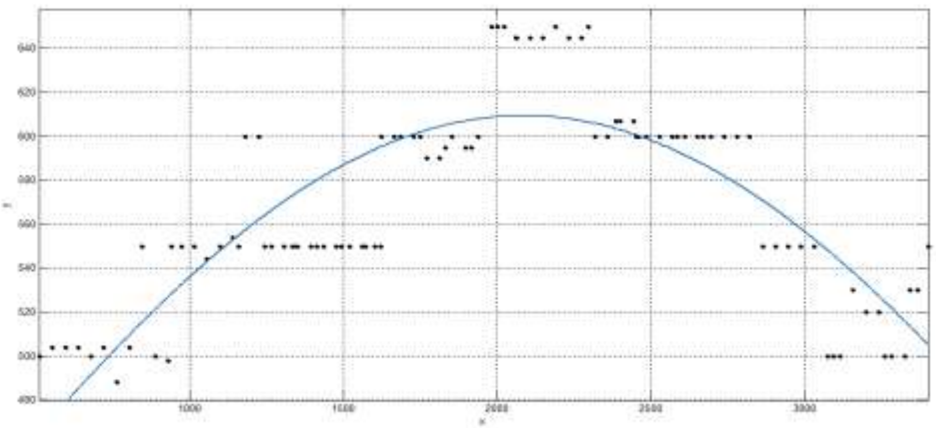


Figure 49: Gaussian Curve Fitting Method, n=1 (measured depth vs fluid flow)

n=2

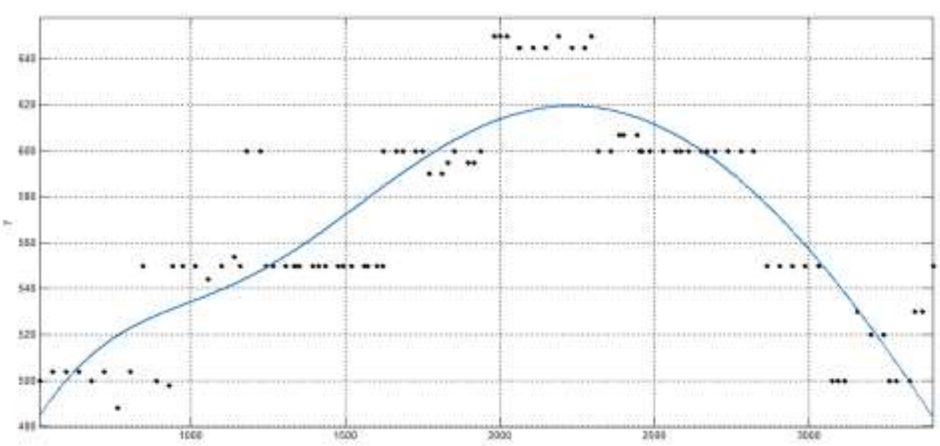


Figure 50: Gaussian Curve Fitting Method, n=2 (measured depth vs fluid flow)

Gaussian method also is not suitable to be used because it does not show the trend of the data with a large error.

5.3.2.3 Fourier Curve Fitting Method

n=1

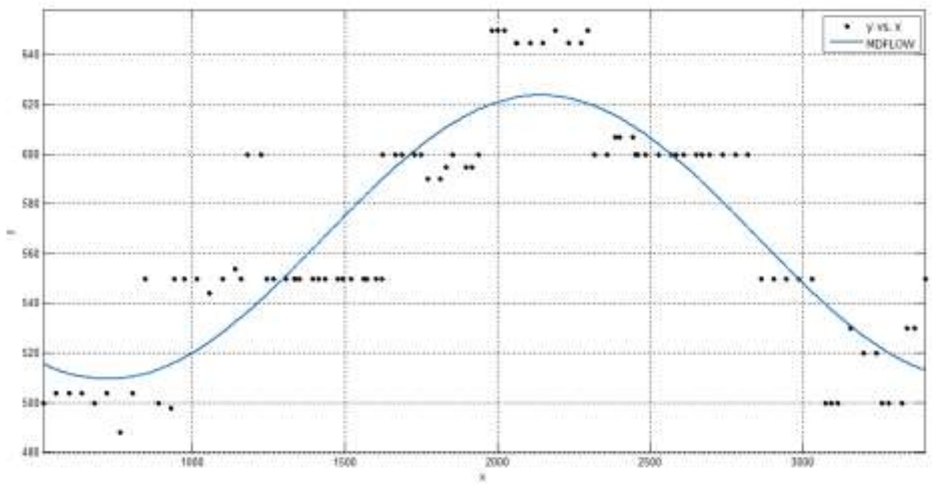


Figure 51: Fourier Curve Fitting Method, n=1 (measured depth vs fluid flow)

n=2

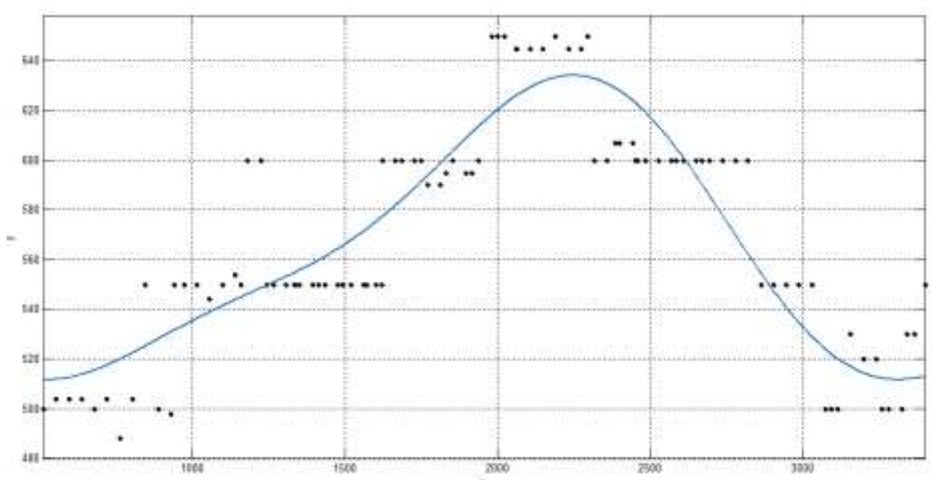


Figure 52: Fourier Curve Fitting Method, n=2 (measured depth vs fluid flow)

Fourier method is not suitable to be used for this data because it is resulting in a higher error.

5.3.2.4 Sine Curve Fitting Method

n=1

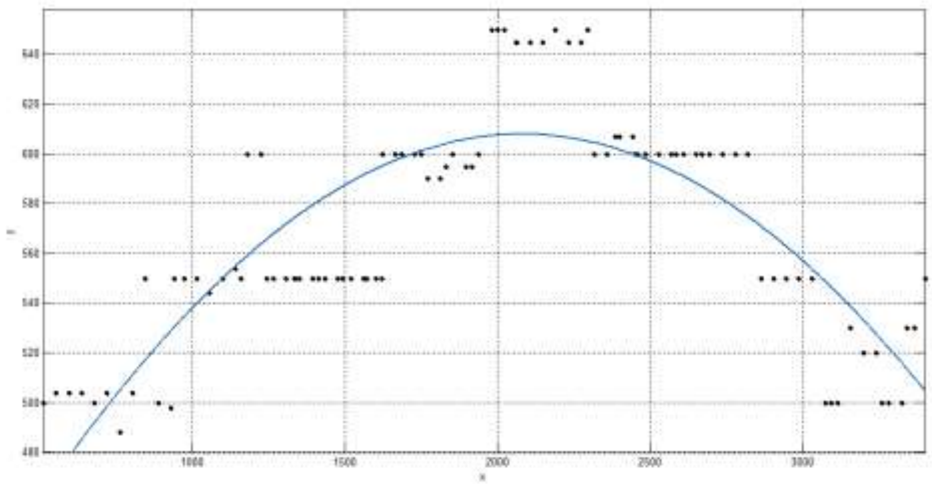


Figure 53: Sine Curve Fitting Method, n=1 (measured depth vs fluid flow)

n=2

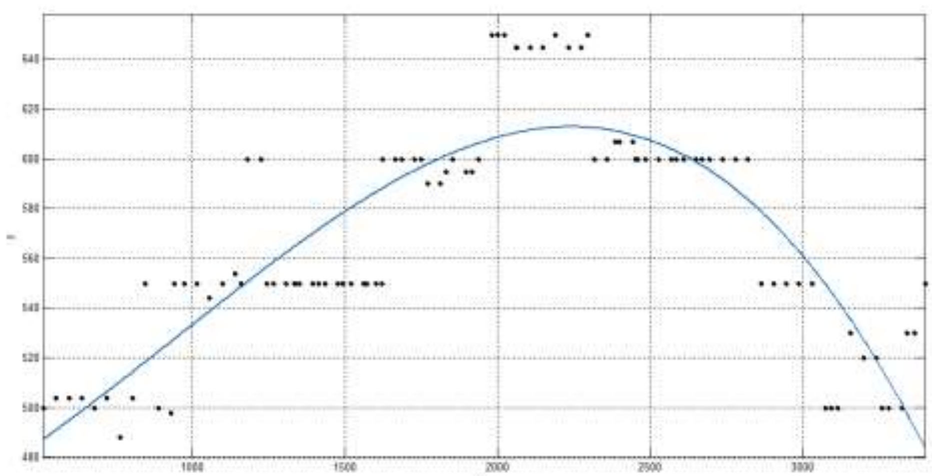


Figure 54: Sine Curve Fitting Method, n=2 (measured depth vs fluid flow)

Sine method is not suitable for this type of data because it does not show the trend of the original data and gives the high error.

5.3.2.5 Smoothing Spline Curve Fitting Method

P=0.85

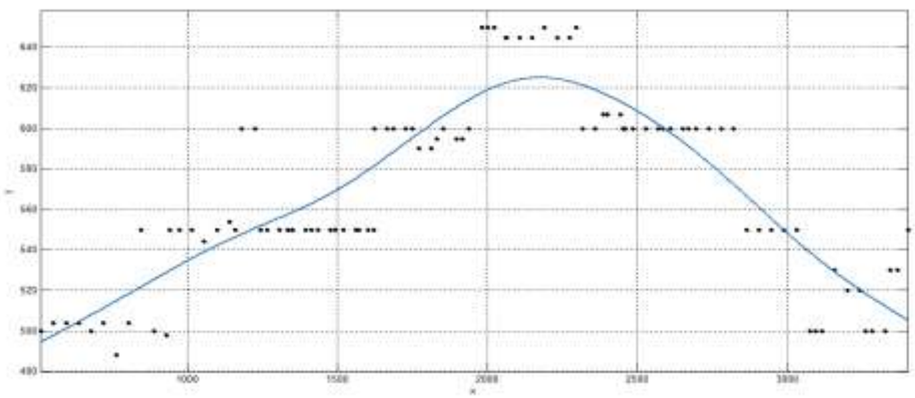


Figure 55: Smoothing Spline Curve Fitting Method, p=0.85 (measured depth vs fluid flow)

P=0.9

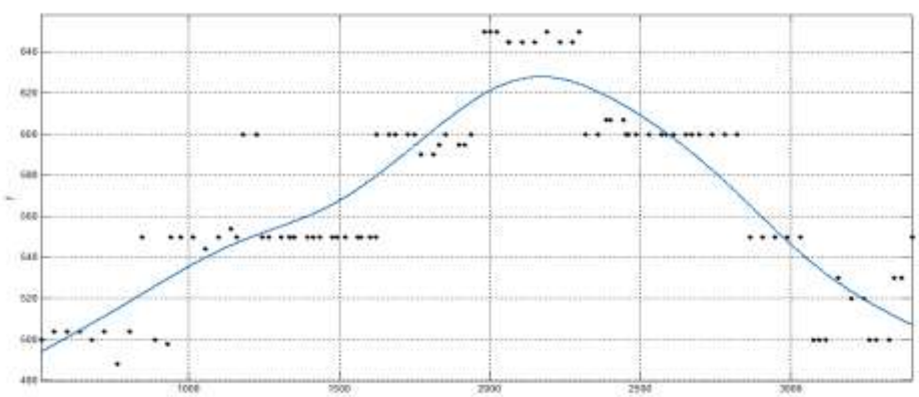


Figure 56: Smoothing Spline Curve Fitting Method, p=0.9

(measured depth vs fluid flow)

P=0.95

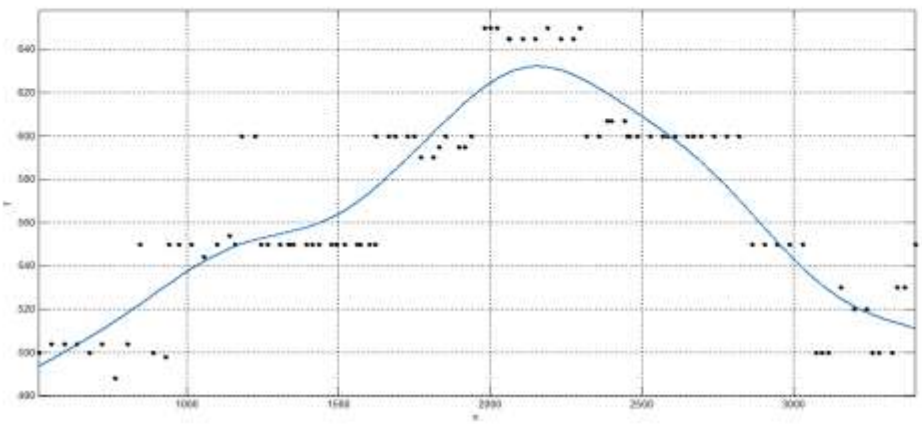


Figure 57: Smoothing Spline Curve Fitting Method, p=0.95 (measured depth vs fluid flow)

P=0.99

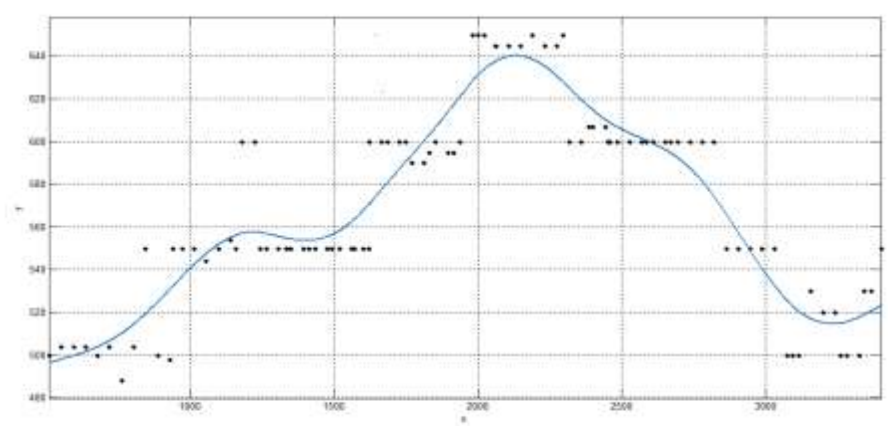


Figure 58: Smoothing Spline Curve Fitting Method, p=0.99 (measured depth vs fluid flow)

Smoothing Spline method is not the best fit for this type of data because it not fully shown the original trend of the data even the value of p is closer to 1. Curve fitting method is not really suitable for this type of data compared to the interpolation method. So new curve fitting method is needed to be applied so that it can show the trend of the data without destroying the original feature of the data.

5.4 Data fitting by using new curve fitting method

From the analysis and research from the interpolation and curve fitting part, it shown that interpolation method is more suitable to be used for fitting purposed because it shown the best fit and not destroying the trend of data. As stated from the literature review part of the new curve fitting analysis, we will combined two type of interpolation method which are PCHIP and Cubic Spline with a several segment division and possible of data points reduction with minimum error within the given tolerance

There are 5 type of results that will be shown in this part which are the original interpolation of PCHIP and Spline, the interpolation of PCHIP and Spline after divided into several segments and some possible point removal according to the algorithm with a minimum errors.

5.4.1 Measured Depth vs Torque

For this data, about 29 data points has been reduce from 44 points. It is about 34% of data reduction has been made with a 14 segments from the total of 43 segments. The new curve fitting result has been shown in the next section.

5.4.1.1Original PCHIP Interpolation

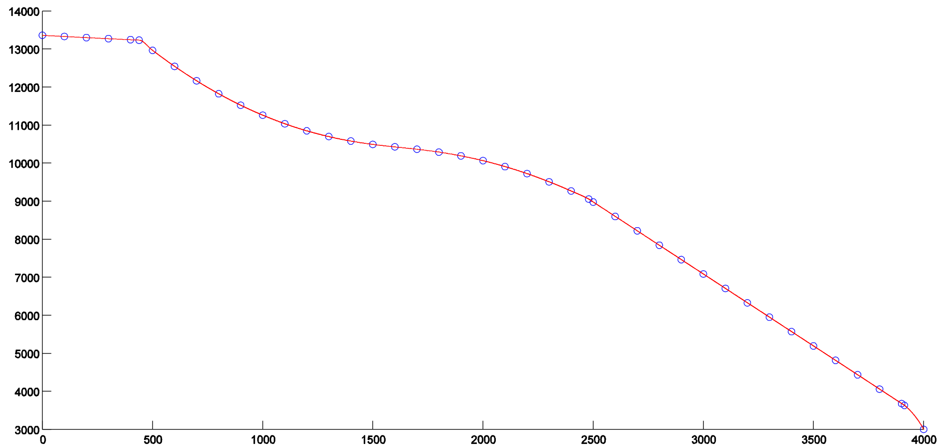


Figure 59: Original PCHIP Interpolation

5.4.1.2 Original Cubic Spline Interpolation

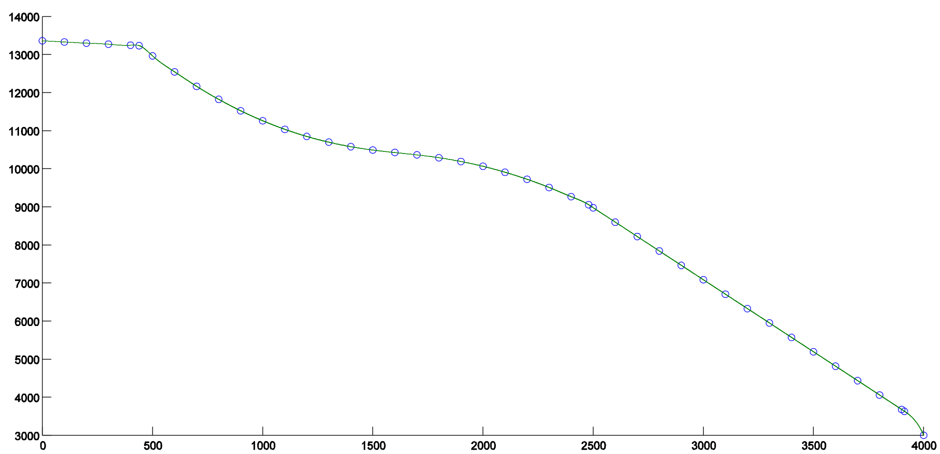


Figure 60: Original Cubic Spline Interpolation, Measured Depth vs Torque

5.4.1.3 PCHIP Interpolation (after divided into several segment and points removal)

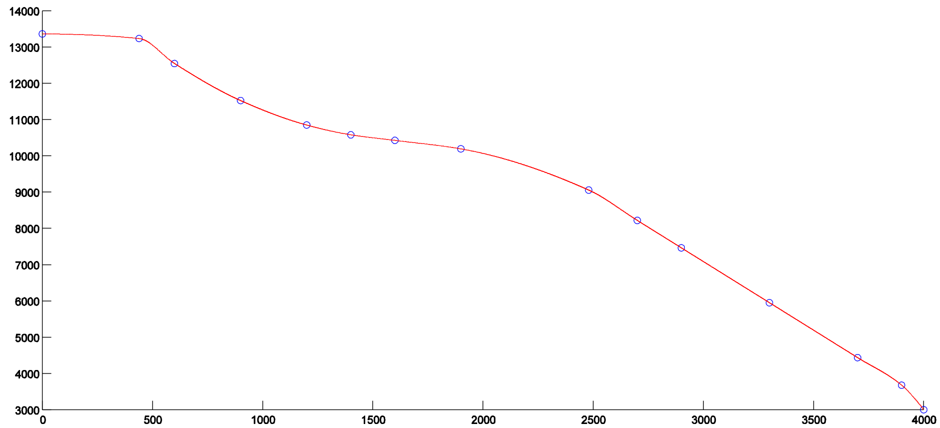


Figure 61:PCHIP Interpolation (after divided into several segment and points removal), Measured Depth vs Torque

5.4.1.4 Cubic Spline Interpolation (after divided into several segment and points removal)

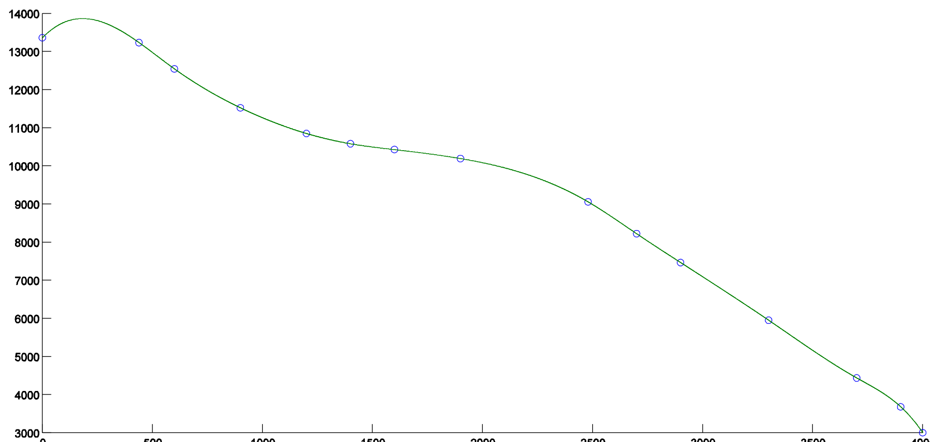


Figure 62:Cubic Spline Interpolation (after divided into several segment and points removal), Measured Depth vs Torque

5.4.1.5 New Curve Fitting Method

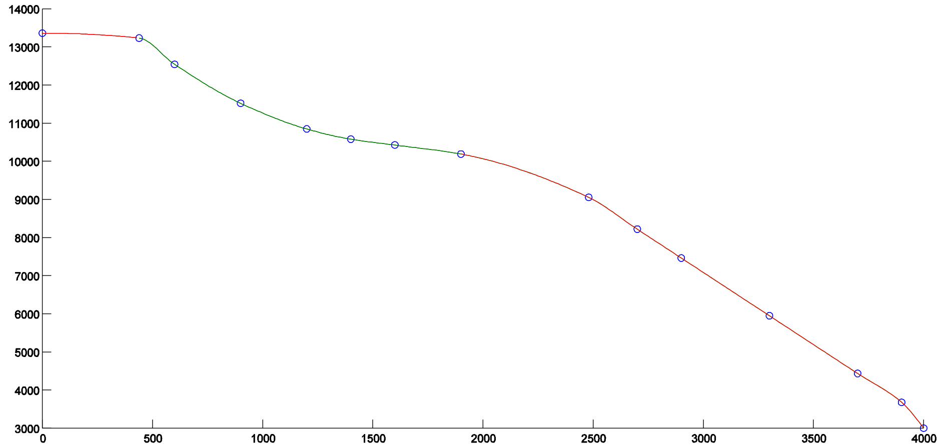


Figure 63: New Curve Fitting Method, Measured Depth vs Torque

Table 10:Measured Depth vs Torque (New Curve Fitting Method)

x-value	original data	PCHIP	Cubic Spline	Error PCHIP	Error Cubic Spline	New Method
0.00	13357.36	13357.36	13357.36	0.0000	0.0000	0.0000
100.00	13326.66	13352.76	13766.59	26.1066	439.9325	26.1066
200.00	13297.15	13325.99	13857.64	38.8354	560.4913	28.8354
300.00	13268.84	13295.48	13705.59	36.6391	436.7487	26.6391
400.00	13241.73	13259.75	13385.50	18.0243	143.7746	18.0243
439.00	13231.33	13231.33	13231.33	0.0000	0.0000	0.0000
500.00	12961.37	13054.47	12972.44	93.1016	11.0718	11.0718
600.00	12541.49	12541.49	12541.49	0.0000	0.0000	0.0000
700.00	12161.50	12169.84	12153.93	8.3416	7.5666	7.5666
800.00	11821.15	11823.10	11815.98	1.9517	5.1657	5.1657
900.00	11520.06	11520.06	11520.06	0.0000	0.0000	0.0000
1000.00	11257.75	11264.35	11260.00	6.6005	2.2464	2.2464
1100.00	11033.55	11036.57	11035.22	3.0176	1.6699	1.6699
1200.00	10846.56	10846.56	10846.56	0.0000	0.0000	0.0000
1300.00	10695.44	10696.79	10694.48	1.3487	0.9584	0.9584
1400.00	10578.09	10578.09	10578.09	0.0000	0.0000	0.0000
1500.00	10490.60	10497.00	10493.28	6.4046	2.6863	2.6863
1600.00	10424.58	10424.58	10424.58	0.0000	0.0000	0.0000
1700.00	10363.52	10353.72	10355.80	9.7947	7.7136	7.7136

1800.00	10288.68	10281.73	10279.33	6.9499	9.3434	9.3434
1900.00	10189.69	10189.69	10189.69	0.0000	0.0000	0.0000
2000.00	10062.69	10066.75	10081.00	4.0625	18.3103	4.0625
2100.00	9906.51	9909.73	9945.82	3.2172	39.3073	3.2172
2200.00	9721.09	9722.13	9776.29	1.0436	55.2010	1.0436
2300.00	9506.89	9506.42	9564.57	0.4731	57.6763	0.4731
2400.00	9264.68	9265.04	9302.81	0.3645	38.1304	0.3645
2479.56	9053.59	9053.59	9053.59	0.0000	0.0000	0.0000
2500.00	8976.09	8997.33	8983.23	21.2360	7.1369	21.2360
2600.00	8597.14	8627.98	8610.77	30.8422	13.6305	30.8422
2700.00	8218.38	8218.38	8218.38	0.0000	0.0000	0.0000
2800.00	7839.79	7843.57	7835.36	3.7792	4.4269	3.7792
2900.00	7461.34	7461.34	7461.34	0.0000	0.0000	0.0000
3000.00	7083.01	7086.80	7089.29	3.7927	6.2870	3.7927
3100.00	6704.75	6708.57	6715.25	3.8172	10.5005	3.8172
3200.00	6326.55	6330.38	6336.02	3.8247	9.4645	3.8247
3300.00	5948.38	5948.38	5948.38	0.0000	0.0000	0.0000
3400.00	5570.19	5573.93	5552.04	3.7380	18.1537	3.7380
3500.00	5191.97	5195.72	5158.28	3.7432	33.6888	3.7432
3600.00	4813.69	4817.46	4781.31	3.7679	32.3789	3.7679
3700.00	4435.31	4435.31	4435.31	0.0025	0.0025	0.0025
3800.00	4056.81	4081.03	4107.62	24.2159	50.8079	24.2159
3900.00	3678.16	3678.16	3678.16	0.0031	0.0031	0.0031
3913.00	3628.93	3584.23	3607.19	44.7031	21.7365	44.7031
4000.00	3000.00	3000.00	3000.00	0.0000	0.0000	0.0000
					Total Errors	304.6541

By using equation (71), we calculate the error by using Root Mean Square Error (RMSE) method.

$$\text{RMSE}=\sqrt{\frac{(304.6541)^2}{44}} = 45.93$$

Comparison between new curve fitting methods with existing method

Table 11: Comparison of RMSE (measured depth vs torque)

Type of Curve Fitting	RMSE
Polynomial (n=9)	48.23
Gaussian (n=2)	82.92
Fourier (n=2)	172.1
Sine (n=2)	280.4
Smoothing Spline (p = 0.95)	159.3
New Curve Fitting Method	45.93

From Table 11, we can see clearly that our new curve fitting method give minimum error within the given tolerance. With the data reduction about 34% respectively, the final fitting curve are acceptable and comparable with the interpolating curves by using cubic spline and PCHIP.

5.4.2 Measured Depth vs Fluid Flow

For this data, about 22 data points has been reduce from 48 points. It is about 54% of data reduction has been made with a 26 segments from the total of 47 segments. The new curve fitting result has been shown in the next section.

5.4.2.1 Original PCHIP Interpolation

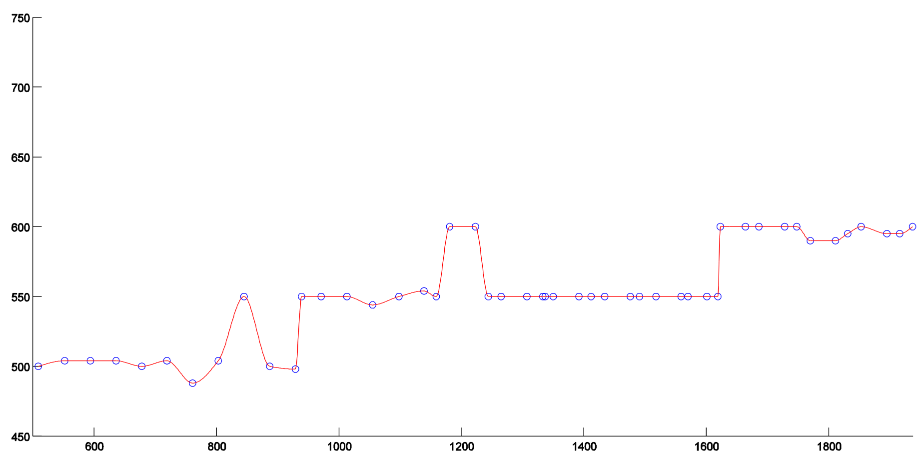


Figure 64: Original PCHIP Interpolation, Measured Depth vs Fluid Flow

5.4.2.2 Original Cubic Splines Interpolation

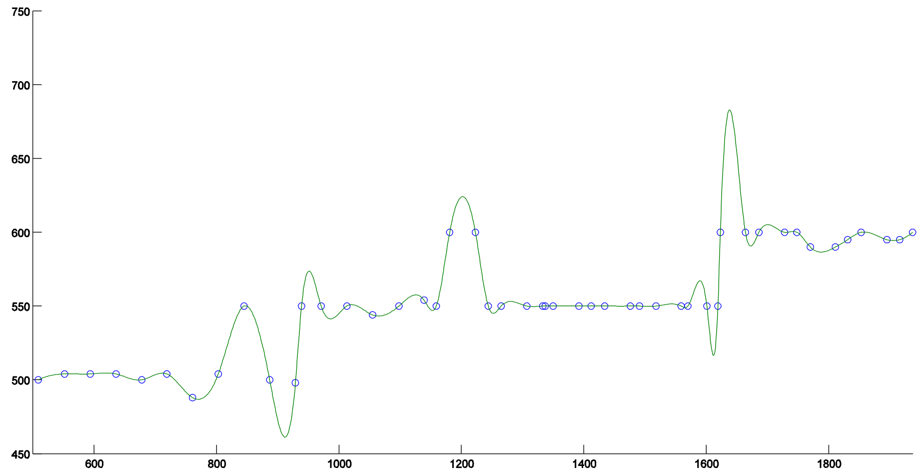


Figure 65: Original Cubic Splines Interpolation, Measured Depth vs Fluid Flow

5.4.2.3 PCHIP Interpolation (after divided into several segment and points removal)

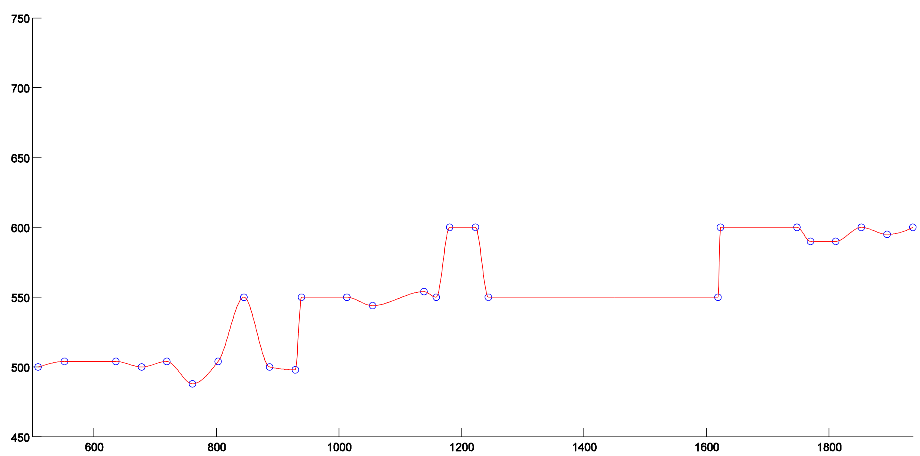


Figure 66:PCHIP Interpolation (after divided into several segment and points removal), Measured Depth vs Fluid Flow

5.4.2.4 Cubic Spline Interpolation (after divided into several segment and points removal)

Figure 67: Cubic Spline Interpolation (after divided into several segment and points removal), Measured Depth vs Fluid Flow

5.4.2.5 New curve fitting method

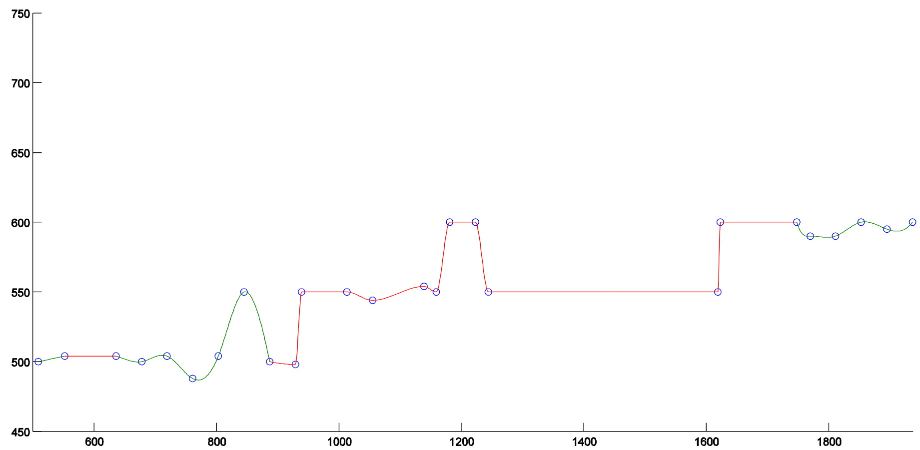


Figure 68: New Curve Fitting Method, Measured Depth vs Fluid Flow

Table 12: Measured Depth vs Fluid FLow (New Curve FItting Method)

x-value	Original Data	PCHIP	Cubic Spline	Error PCHIP	Error Cubic Spline	New Method
509	500	500.0000	500.0000	0.0000	0.0000	0.0000
552	504	504.0000	504.0000	0.0000	0.0000	0.0000
594	504	504.0000	506.3445	0.0000	2.3445	0.0000
636	504	504.0000	504.0000	0.0000	0.0000	0.0000
678	500	500.0000	500.0000	0.0000	0.0000	0.0000
719	504	504.0000	504.0000	0.0000	0.0000	0.0000
761	488	488.0000	488.0000	0.0000	0.0000	0.0000
803	504	504.0000	504.0000	0.0000	0.0000	0.0000
845	550	550.0000	550.0000	0.0000	0.0000	0.0000
887	500	500.0000	500.0000	0.0000	0.0000	0.0000
929	498	498.0000	498.0000	0.0000	0.0000	0.0000
939	550	550.0000	550.0000	0.0000	0.0000	0.0000
971	550	550.0000	610.4041	0.0000	60.4041	0.0000
1013	550	550.0000	550.0000	0.0000	0.0000	0.0000
1055	544	544.0000	544.0000	0.0000	0.0000	0.0000
1098	550	549.1785	561.1267	0.8215	11.1267	0.8215
1139	554	554.0000	554.0000	0.0000	0.0000	0.0000
1159	550	550.0000	550.0000	0.0000	0.0000	0.0000
1181	600	600.0000	600.0000	0.0000	0.0000	0.0000
1223	600	600.0000	600.0000	0.0000	0.0000	0.0000
1244	550	550.0000	550.0000	0.0000	0.0000	0.0000
1265	550	550.0000	482.3238	0.0000	67.6762	0.0000
1307	550	550.0000	314.0345	0.0000	235.9655	0.0000
1333	550	550.0000	198.7168	0.0000	351.2832	0.0000
1337	550	550.0000	180.8164	0.0000	369.1836	0.0000
1350	550	550.0000	123.0034	0.0000	426.9966	0.0000
1392	550	550.0000	-50.5665	0.0000	600.5665	0.0000
1412	550	550.0000	-119.7034	0.0000	669.7034	0.0000
1434	550	550.0000	-180.3215	0.0000	730.3215	0.0000
1476	550	550.0000	-235.5557	0.0000	785.5557	0.0000
1491	550	550.0000	-231.3755	0.0000	781.3755	0.0000
1518	550	550.0000	-185.5634	0.0000	735.5634	0.0000
1559	550	550.0000	-5.8804	0.0000	555.8804	0.0000
1570	550	550.0000	68.0407	0.0000	481.9593	0.0000
1601	550	550.0000	342.3579	0.0000	207.6421	0.0000
1619	550	550.0000	550.0000	0.0000	0.0000	0.0000
1623	600	600.0000	600.0000	0.0000	0.0000	0.0000
1664	600	600.0000	837.4172	0.0000	237.4172	0.0000

1686	600	600.0000	810.9497	0.0000	210.9497	0.0000
1728	600	600.0000	653.5586	0.0000	53.5586	0.0000
1748	600	600.0000	600.0000	0.0000	0.0000	0.0000
1770	590	590.0000	590.0000	0.0000	0.0000	0.0000
1811	590	590.0000	590.0000	0.0000	0.0000	0.0000
1831	595	594.6431	594.9975	0.3569	0.0025	0.0025
1853	600	600.0000	600.0000	0.0000	0.0000	0.0000
1895	595	595.0000	595.0000	0.0000	0.0000	0.0000
1916	595	596.2500	593.6398	1.2500	1.3602	1.3602
1937	600	600.0000	600.0000	0.0000	0.0000	0.0000

$$\text{RMSE} = \sqrt{\frac{(2.1842)^2}{48}} = 0.3152$$

Table 13 : Comparison of RMSE (measured depth vs fluid flow)

Type of Curve Fitting	**RMSE**
Polynomial (n=9)	17.55
Gaussian (n=2)	22.65
Fourier (n=2)	19.76
Sine (n=2)	24.13
Smoothing Spline (p = 0.99)	16.04
New Curve Fitting Method	0.3152

From Table 13, our new curve fitting also give less error compared to the other fitting methods by reduction of 54% of data (26 data points were removed) and 22 segments from the total of 47 segments. Clearly our new curve fitting methods give the best results compared with the existing methods.

5.4.3 Temperature Correction Factor for Nitrogen Gas (Typical Gas Lift)

In this section we apply our new curve fitting method to find the best fitting by using the data taken from the Equation (72). This is the uniformly type of data because the value of y-axis changed according to the function. Equation (72) is the original function to calculate the value of Ct (y-axis).

$$Ct = 1/[1.0\ +\ 2.237\ x\ (Tv - 60)] \tag{72}$$

x-value = Tv, y-value = Ct

This data plot for the original function can be seen in in Winkler et al (1989). About 21 of data reduction from 25 data points and by using only 4 segment from the total of 25 segment has been analyzed to give the best result for our new curve fitting. The result has been shown in the next section.

5.4.3.1 Original PCHIP Interpolation

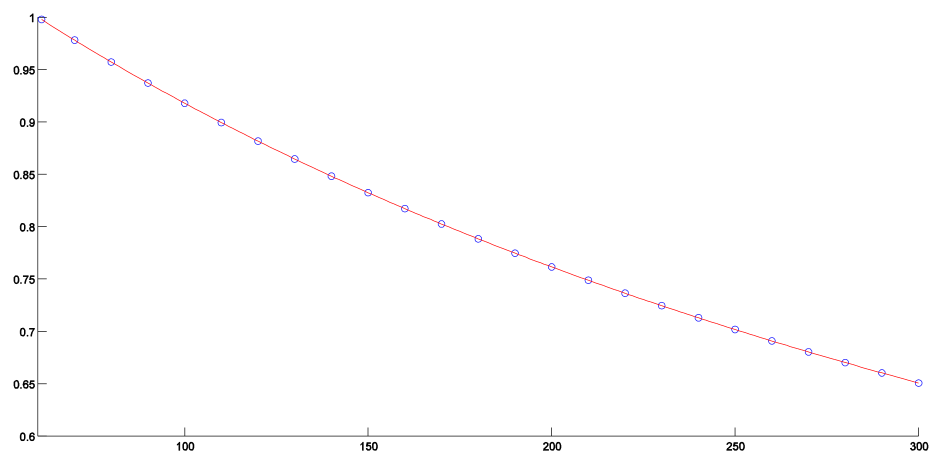


Figure 69: Original PCHIP Interpolation, Temperature Correction Factor

5.4.3.2 Original Cubic Spline interpolation

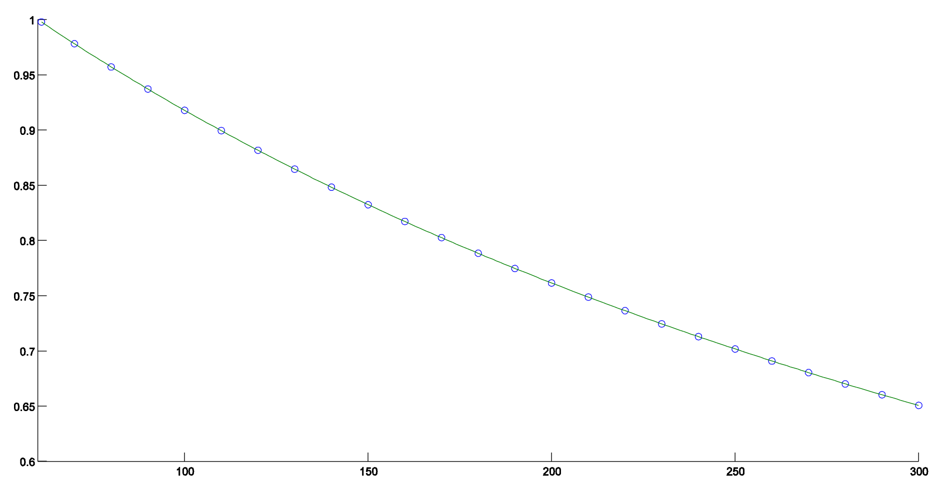


Figure 70:Original Cubic Spline interpolation, Temperature Correction Factor

5.4.3.3 PCHIP Interpolation (after divided into several segment and points removal)

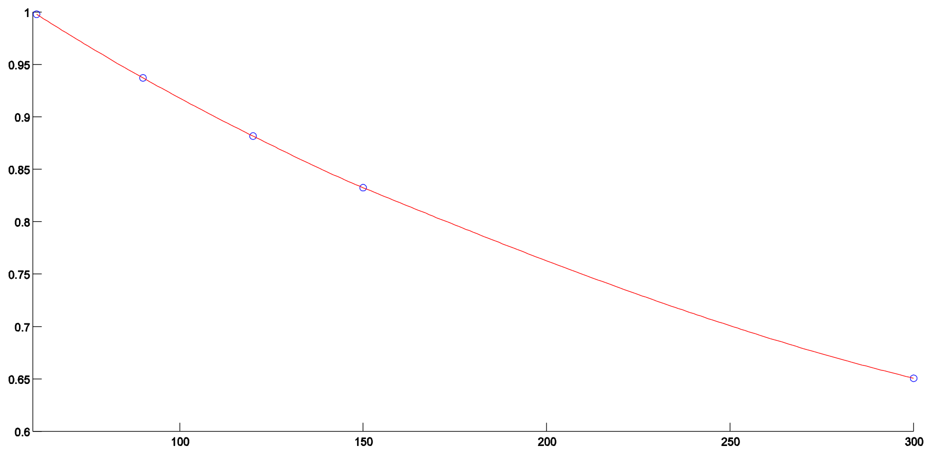


Figure 71: PCHIP Interpolation (after divided into several segment and points removal), Temperature Correction Factor

5.4.3.4 Cubic Spline Interpolation (after divided into several segment and points removal)

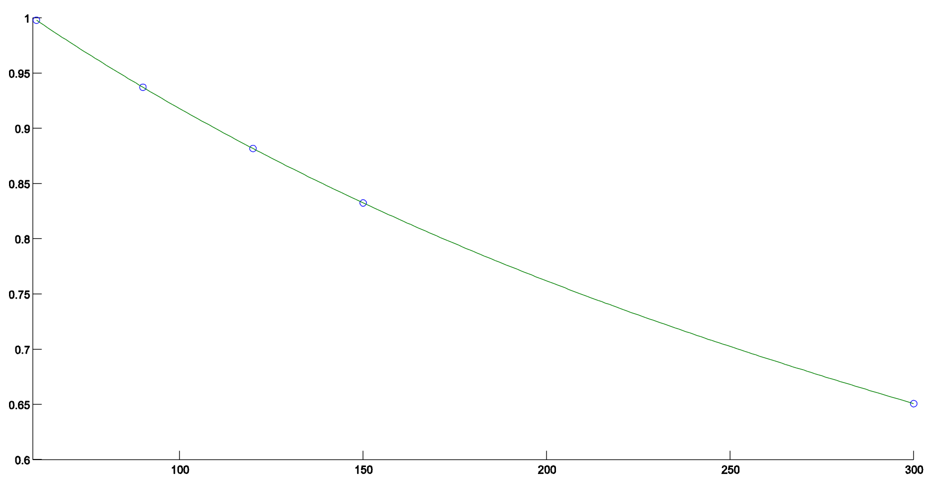


Figure 72: Cubic Spline Interpolation (after divided into several segment and points removal), Temperature Correction Factor

5.4.3.5 New curve fitting method

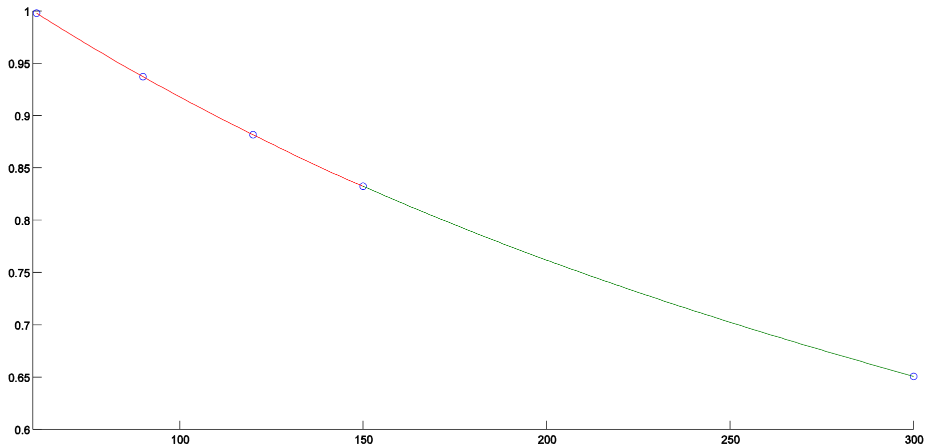


Figure 73:New curve fitting method, Temperature Correction Factor

Table 14: Temperature Correction Factor (New Curve Fitting Method)

Temperature	**Original Function**	**Cubic Spline**	**PCHIP**	**Error CS**	**Error PCHIP**	**New Method**
61	0.9978	0.9978	0.9978	0.0000	0.0000	0.0000
70	0.9781	0.9781	0.9782	0.0000	0.0000	0.0000
80	0.9572	0.9572	0.9572	0.0000	0.0000	0.0000
90	0.9371	0.9371	0.9371	0.0000	0.0000	0.0000
100	0.9179	0.9179	0.9179	0.0000	0.0000	0.0000
110	0.8994	0.8994	0.8994	0.0000	0.0000	0.0000
120	0.8817	0.8817	0.8817	0.0000	0.0000	0.0000
130	0.8646	0.8646	0.8646	0.0000	0.0000	0.0000
140	0.8482	0.8482	0.8481	0.0000	0.0001	0.0001
150	0.8324	0.8324	0.8324	0.0000	0.0000	0.0000
160	0.8172	0.8172	0.8181	0.0000	0.0009	0.0000
170	0.8025	0.8026	0.8039	0.0001	0.0014	0.0001
180	0.7884	0.7885	0.7899	0.0001	0.0015	0.0001
190	0.7747	0.7749	0.7762	0.0002	0.0015	0.0002

200	0.7615	0.7617	0.7627	0.0002	0.0012	0.0002
210	0.7487	0.7490	0.7496	0.0003	0.0009	0.0003
220	0.7364	0.7369	0.7367	0.0005	0.0003	0.0005
230	0.7245	0.725	0.7243	0.0005	0.0002	0.0005
240	0.7129	0.7135	0.7123	0.0006	0.0006	0.0006
250	0.7017	0.7024	0.7007	0.0007	0.0010	0.0007
260	0.6909	0.6915	0.6896	0.0006	0.0013	0.0006
270	0.6804	0.681	0.679	0.0006	0.0014	0.0006
280	0.6702	0.6707	0.669	0.0005	0.0012	0.0005
290	0.6603	0.6606	0.6595	0.0003	0.0008	0.0003
300	0.6507	0.6507	0.6507	0.0000	0.0000	0.0000
			Total Error	**0.0057**	**0.0151**	**0.0053**

From Table 14, new curve fit the interpolation data very well with data reduction of 21. The total error is smaller compare with cubic spline and PCHIP. Thus we conclude that in order to fit the petroleum engineering data, the engineer may use our new curve fitting methods.

From all the results that have been collected, we can summarize the advantages of our new curve fitting method which are:

1. It has fewer segments of fitting curves compare to the existing scheme. For example, from Table 12, our new curve method only has 26 data points compared to the original segments which is 49 data points.
2. Our curve fitting method has data reduction capability. By having data reduction, we may reduce the storage to store the data. From Table 12, in total there are 54% data reduction which is quite good.
3. Our new curve fitting method algorithm can be used for other applications
4. Engineer may use the new curve fitting method for interpolation and approximation purposes since its provides fitting curves with minimum error and it is the best compared to the other existing schemes.

CHAPTER 6

CONCLUSION AND RECOMMENDATION

6.1 Conclusion

The main topic addresses in this research is data interpolation and data approximation (fitting) by using spline interpolation and curve fitting methods. Due to the fact that the existing curve fitting methods have some weaknesses, we propose new curve fitting method to overcome those problems. Our new curves fitting methods give the best results compared with existing methods. One of the main advantages of our new curve fitting methods is that it has the capability of data reduction. By having data reduction the engineer may reconstruct the curves or even surfaces with few data points but maintain higher accuracy of the fitting curves or surfaces. In fact data reduction is crucial when the user require to fitting the thousands of data sets.

6.2 Recommendation

For future work recommendation, our new curve fitting method can be used in image compression since our main strategy consists of data reduction which is basic concept in data compression. Besides, our new curve fitting method can be extended for surfaces problems. This will help the engineer to be able to visualized 3 Dimensional (3D) data. In this case, we may reduce the data very significantly. Thus this will increase the compression ratio (CR).

REFERENCES

1. Karim, S.A.A., B.A. Karim, M.K. Hasan andJ. Sulaiman. Font Designing using Generalized Ball Basis. EnCon2010, 3rd Engineering Conference on Advancement in Mechanical and Manufacturing for Sustainable Environment, April 14-16, 2010, Kuching, Sarawak, Malaysia.

2. Karim, S.A.A. and Kong, V.P. (2011). Gaussian Scale-Space and Discrete Wavelet Transform For Data Smoothing. In International Conference on Electrical, Control and Computer Engineering (INCECCE), Pahang, Malaysia, June 21-22, 2011, Universiti Malaysia Pahang and IEEE Computer Society. pp. 344-348.

3. Karim, S.A. A. and Yahya, N. (2013). Seabed Loging Data Curve Fitting using cubic Splines. Applied Mathematical Sciences. Vol. 7, 2013, no. 81, 4015 - 4026,

4. Karim, S.A.A. and Singh, B.S.M. (2013). Global Solar Radiation Modeling using Polynomial Fitting. Applied Mathematical Sciences, Vol. 8, 2014, no. 8, 367 - 378.

5. Karim, S.A.A., Hashim, A., Bakar, N.A.A. and Hasni, R. (2008). Rational Bézier Curves and Surfaces and its Application in Computer Aided Geometric Design (CAGD). In Proceedings of the Sixteenth National Symposium on Mathematical Sciences, 3-5 Jun 2008 at Hotel Renaissance, Kota Bharu, Malaysia.

6. Zamani, Mehdi. (2012). An Applied Two-Dimensional B-spline Model For Interpolation of Data. Vol, Issue 2, July-December (2012).pp.322-336

7. Zamani, Mehdi (2009). Three Simple Spline Methods for Approximation and Interpolation of Data. Contemporary Engineering Sciences, Vol 2, 2009, no. 8, 373-381.

9. IIk D, Anderson D.M. and Valko P.P. (2006) Analysis of Gas Well Reservoir Performance Data Using B-spline Deconvolution. SPE Gas Technology Symposium, May 15-17, 2006, Calgary, Alberta, Canada.

10. Zamani Mehdi. (2011). Approximation by a kind of B-spline method. Indian Journal of Science and Technology. Vol 4, February 2011. No.2, pp. 321-311

11. Salomon, David (2005). Curves and Surfaces for Computer Graphics. Springer Verlag, August 2005.

12. Bradie, Brian (2006). A Friendly Introduction to Numerical Analysis. Pearson International Edition.

13. Komzik, Louis (2006). Approximation Techniques for Engineers. CRC Press,2006, 296 p.

14. Lee, T. Kravaris, C. and Seinfield, J.H. (1986). History Matching by Spline Approximation and Regularization in Single Phase Reservoir. Paper SPE 13931, SPE Reservoir Engineering.

15. Ahmadi M. (2012). Improving Well-Performance-Data Analysis in Laplace Space by using Cubic Splines and Boundary Mirroring. Paper SPE 159734. SPE Reservoir Engineering.

16. Won, Y., Wenyu, C., Tae-Sang, C., & Morris, J. (2005). Applied numerical methods using matlab. John Wiley & Sons, Inc.

17. Singiresu, S. R. (2002). Applied numerical methods for engineers and scientists (M. J. Horton, Ed.). Tom Robbins Prentice Hall, Inc. Upper Saddle River, New Jersey 07458.

18. Kruger, C. J. C. (n.d.). Constrained cubic spline interpolation for chemical engineering applications.

19. Addendum. (n.d.). Interpolation polynomial. (Spline Method of Interpolation)

20. Winkler, H.W., and Ends, P.T., 1989, “Algorithm for more accuretaly predicting nitrogen-charged gas lift valve operation at high pressure and temperatures”, SPE Production Operations Symposium, SPE 18871, 415-422, March.

PUBLICATIONS

1. Karim, S.A.A., Rosli, M.A. and Mustafa, M.I. (2014). Cubic spline interpolation for petroleum engineering data, Applied Mathematical Sciences, Vol. 8, 2014, no. 102, 5083-5098. Scopus and Q3.